

Preparing for disaster

Earth scientists should find better mechanisms to disseminate facts about the risks of natural disasters, to help local populations make the necessary preparations.

These days, science offers ever-more-comprehensive assessment of the risks of earthquakes, volcanic eruptions, storms and floods, and technology offers more sophisticated approaches for coping with them. Yet growing urban populations — as well as large rural populations in places such as northern Pakistan and Kashmir, now suffering the fallout from October's massive earthquake — remain hugely vulnerable to such disasters (see page 903).

There is plenty of evidence that the right combination of scientific knowledge, experience, planning and common sense can substantially reduce the risks posed by natural disasters. One such example pertains to Hilo in Hawaii, which was badly damaged by a tsunami in 1946. As a consequence, scientific research into the causes and the physical behaviour of the giant waves was intensified, leading in 1949 to the creation of the Pacific Tsunami Warning Center. The effectiveness of the system was put to the test in 1960, when another tsunami flooded the city. Thanks to building restrictions and regular exercises in preparedness and emergency behaviour, Hilo has again become a relatively safe place to live.

Unfortunately, preparation for a tsunami in the Pacific is the exception, rather than the rule. The dozen countries that were affected by the deadly Indian Ocean tsunami a year ago had made few preparations, despite scientists' familiarity with the risks of such an event. Many countries around the world, such as Turkey and Iran, remain unable or unwilling to take the necessary steps to prepare for disasters that specialists believe are waiting to happen.

Scientists who study these risks have a critical and valuable role to play in ensuring that every effort is made to raise public and political awareness of impending risks. The effective communication of risk is a non-trivial problem: individual researchers who study fault ruptures, volcanoes or cyclone thermodynamics are not always well

positioned to publicize their findings widely, and one cannot always expect local policy-makers and planners to delve directly into the scientific literature for information. So imaginative approaches are needed to forge effective links between the two groups.

Some effort is now being made to implement such approaches at the global level. For example, the World Conference on Disaster Reduction, held last January in Kobe, Japan, called for a worldwide risk-management strategy coordinated by the United Nations. Such a strategy needs solid scientific support, and David King, science adviser to the British government, has suggested setting up an International Science Panel for Natural Hazard Assessment to provide it. A proposed joint initiative by the United Nations' Development Programme and the World Bank might fulfil the same purpose without the need to establish a new organization.

No amount of international coordination activity will make much difference, however, in regions where poverty, illiteracy and corruption stymie preparations against disaster. In many parts of the world, compliance with regulations to ensure that buildings are constructed to withstand earthquakes, for example, would be totally beyond the means of the local population. From Tehran to New Orleans, disaster reduction has an immense social dimension — people can be protected only as part of a broader fight against poverty.

That said, risk management can be improved through international mechanisms that will feed the best science to decision-makers. Global thinking is vital — but saving lives ultimately requires preparation at a local level. ■

"Disaster reduction has an immense social dimension — people can be protected only as part of a broader fight against poverty."

Europe's right stuff

The European Space Agency is making good use of its funds for space exploration.

When ministers from the member states of the European Space Agency (ESA) met in Berlin last week, they made a number of good decisions. They unexpectedly agreed to provide the agency with all of the 2.5% funding boost that it had requested for science missions. They also agreed to funds for additional missions for Mars surface exploration and Earth monitoring. All of this is welcome. Even more surprising, and equally welcome, is their decision not to spend money on the development of a new Russian-led space-plane, Clipper — a decision that surprised some observers but that should be seen as a smart move. Europe does

not need the ability to launch humans into space, and should resist further attempts by Russia to solicit funds for it.

Whether Clipper can actually work is an open question. The history of small space-plane programmes is long and unhappy, with ESA's abortive Hermes programme being as big a let-down as all the others. Many believe that there are fundamental flaws in the idea of adding to the mass of spacecraft by giving them wings with which to fly, rather than just settling for a controlled plummet in the manner of Russia's Soyuz capsules and the United States' proposed Crew Exploration Vehicle. But even if Clipper stood a realistic chance of working, its development would be a hugely inappropriate use of European taxpayers' money.

The idea of human space flight is an inspiring and noble one. Unfortunately, achieving it means devoting vast resources to some markedly unproductive goals in a way normally only possible under political systems that are neither inspiring nor noble. The United



States is the only democracy that has risen to the challenge, and remarkable though its achievement in this sphere has been, it has left an ambivalent legacy.

The Apollo programme was politically sustainable because it resonated with various aspects of America's self-image as a nation that is technologically peerless, internationally exceptional and defined by the notion of frontiers. These resonances persist today. Few Americans are passionately devoted to the space programme, but many think of it, on the occasions they have cause to, with affection. Given the great cost of its limited benefits, this popular support seems a touch perverse.

For any US president, the political cost of being the person who abandons the dream of space flight outweighs the financial cost of "keeping the dream alive" (the term under which this sort of support for the aerospace establishment is invariably masked). At the same time, the financial costs that would have to be borne in order dramatically to expand the role of astronauts in space exploration are seen as outweighing any possible political rewards from such an expansion. So there is a compromise: the United States is left with an extraordinarily expensive and simultaneously rather unambitious

programme, the main purpose of which is its own continuation.

This may not be a very good deal for US taxpayers, but it does have benefits for the world at large. It means that human presence in orbit, a largely symbolic matter, is not restricted to the citizens of one-party states (China) and their cash-strapped successors (Russia, a country which in terms of gross domestic product per capita ranks between Chile and Malaysia). It is good to know that, if there are men and women beyond Earth, some of them should be from democracies. But unless the US political landscape undergoes a radical shift, that will remain the case whatever Europe does, and it is hard to see what extra value is to be gained by any other democracies deciding to join in the venture. There are better ways to convince the world of your technological prowess.

The fact that the United States cannot bring itself to give up human space flight is, at the end of the day, no reason for Europe to join in with a programme of its own. ■

"The fact that the United States cannot bring itself to give up human space flight is no reason for Europe to join in with a programme of its own."

Wiki's wild world

Researchers should read Wikipedia cautiously and amend it enthusiastically.

Sometimes the stupid-sounding ideas turn out to be the ones that take off. Almost five years ago, a free online encyclopaedia known as Wikipedia was launched. To those familiar with the peer-review process, the premise behind the new publication seemed crazy: any user, regardless of expertise, can edit the entries. It sounded like a method for creating garbled and inaccurate articles, and many critics said so.

Fast-forward to 2005, and some of that criticism is looking misplaced. Wikipedia is now a huge reference source, with something approaching a million articles in the English version alone. It's true that many of its entries are confusing and badly structured; some of them are badly wrong, and sometimes the errors are deliberate. After the discovery of an outrageously false description of John Seigenthaler, a former editor of *The Tennessean* newspaper, Wikipedia's publishers introduced registration in an attempt to discourage (though it cannot prevent) "impulsive vandalism".

But as an investigation on page 900 of this issue shows, the accuracy of science in Wikipedia is surprisingly good: the number of errors in a typical Wikipedia science article is not substantially more than in Encyclopaedia Britannica, often considered the gold-standard entry-level reference work. That crazy idea is starting to look anything but stupid.

So can Wikipedia move up a gear and match the quality of rival reference works? Imagine the result if it did: a comprehensive, accurate and up-to-date reference work that can be accessed free from Manhattan to rural Mongolia. To achieve this, Wikipedia's administrators will have to tackle everything from future funding problems — the site is maintained by public donations — to doubts about

whether enough new contributors can be found to increase the quality of the mushrooming number of entries. That latter point is critical, and here scientists can make a difference.

Judging by a survey of *Nature* authors, conducted in parallel with the accuracy investigation, only a small percentage of scientists currently contribute to Wikipedia. Yet when they do, they can make a significant difference. Wikipedia's non-expert contributors are, by and large, dedicated to getting things right on the site. But scientists can bring a critical eye to entries on subjects they study, often highlighting errors and misunderstandings that others have unintentionally introduced. They can also start entries on topics that other users may not want to tackle. It is no surprise, for example, that the entry on 'spin density wave' was originated by a physicist.

Editing pages is not always straightforward, as some users may disagree with changes. In politically sensitive areas such as climate change, researchers have had to do battle with sceptics pushing an editorial line that is out of kilter with mainstream scientific thinking. But this usually requires no more than a little patience. Wikipedia's users are generally interested in the reasoning behind proposed changes to articles. Backing up a claim with a peer-reviewed reference, for example, makes a world of difference.

Nature would like to encourage its readers to help. The idea is not to seek a replacement for established sources such as the Encyclopaedia Britannica, but to push forward the grand experiment that is Wikipedia, and to see how much it can improve. Select a topic close to your work and look it up on Wikipedia. If the entry contains errors or important omissions, dive in and help fix them. It need not take too long. And imagine the pay-off: you could be one of the people who helped turn an apparently stupid idea into a free, high-quality global resource. ■

"Scientists can bring a critical eye to entries on subjects they study, highlighting errors that others have unintentionally introduced."

RESEARCH HIGHLIGHTS

Not so hot spots

Earth Planet. Sci. Lett.

doi:10.1016/j.epsl.2005.10.012 (2005)

The volcanoes of Hawaii (pictured) may give vent to the fury of the deep Earth, but the plumes of magma that feed such hot spots carry little of the core's heat, say Eric Mittelstaedt at the University of Hawaii and Paul Tackley at the Swiss Federal Institute of Technology Zurich.

The researchers' computer simulations of a single plume show that it transports only 10–50% of the heat crossing the boundary between the Earth's core and the mantle in the region where the plume rises. This has been observed in simulations before, but the new work makes more realistic assumptions about the geophysical properties of the mantle, such as the variation in its viscosity.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

G. B. LEWIS/GETTY IMAGES

PHOTONICS**Green light for white light**

Adv. Mat. **17**, 2974–2978 (2005); *Appl. Phys. Lett.* **87**, 231901 (2005)

Getting light-emitting diodes (LEDs) to produce white light has always been difficult — such devices are mostly made by combining red, blue and green emitters. But now there are two ways to make white light from a single emissive ingredient.

Lixiang Wang and his colleagues of the Changchun Institute of Applied Chemistry in China grafted red- and green-emitting organic groups on to the blue-emitting main chain of an organic polymer to create a white-light organic LED.

In a separate study, Koji Nishizuka of Kyoto University and his co-workers in Japan drew white light out of an inorganic semiconductor, indium gallium nitride, by varying the amount of indium through the device's crystalline film: the more indium,

the longer the emission wavelength. The resulting rainbow mixture of wavelengths combines to give a greenish white light.

IMMUNOLOGY**Making mice tweak**

Cell **123**, 931–944 (2005)

A protein known as TWEAK seems to act as a brake on innate immunity, say Avi Ashkenazi and his co-workers at the company Genentech in San Francisco, California.

Innate immunity is the body's first line of defence against infection. It also influences our later, adaptive immune response. Ashkenazi's group bred mice deficient in the *Tweak* (or *Tnfsf12*) gene to investigate its role.

TWEAK-deficient mice died from overstimulation of their innate immune cells (pictured left) on exposure to a bacterial toxin. But these mice also made more of the proteins that can predispose cells to secrete pro-inflammatory signals, which can stimulate adaptive immunity. The researchers suggest that manipulating TWEAK may help in treatments for infections and auto-immune diseases.

CHEMISTRY**Double delivery**

J. Am. Chem. Soc. doi:10.1021/ja056841t (2005)

Micelles are aggregates of molecules that form compartments, which can be useful for packaging and transporting smaller chemicals such as pharmaceuticals.

Pushing these structures to new limits, Timothy Lodge and his colleagues from the University of Minnesota in Minneapolis

have created a micelle with two internal compartments, and have shown that a different type of molecule can be stored in each. This means that two chemicals that would otherwise react with each other could potentially be delivered to the same place at the same time.

The multicompartment micelle is made from a polymer that has three different arms. The polyethylene and polyperfluoropropylene oxide arms segregate within the micelle core to form the distinct storage areas. Polyethylene oxide pokes out of the sides, making the packages soluble in water.

NEUROSCIENCE**Reflecting badly**

Nature Neurosci. doi:10.1038/nn1611 (2005)

Brain cells that let us mirror the actions of others could hold the key to the social problems seen in autism, say researchers.

Mirella Dapretto and her team at the University of California, Los Angeles, studied mirror neurons in normal children and in those with mild autism. These neurons normally fire both when a person performs a particular action and when that person sees someone else performing the same action. They are thought to be key to our ability to automatically understand and empathize with the actions and emotions of others.

The team performed brain scans on both sets of children while they either imitated or just watched pictures of emotional facial expressions. The mirror neurons in the autistic children were far less active than normal, and were least active in the children whose social skills were worst affected.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

GENETICS

Ripped genes

Nature Genet. doi:10.1038/ng1695; 10.1038/ng1696; 10.1038/ng1697 (2005)

Our genomes are shot full of holes, according to three reports from researchers in the United States and Britain. Collectively, they have discovered 1,000 previously unknown gaps, or deletions, in the chromosomes of normal individuals.

The deletions, ranging from tens to hundreds of thousands of DNA base pairs, are smaller than have previously been found. The teams hope that the findings will shed light on how variations in our genomes affect our predisposition to disease.

The researchers made use of data already collected as part of the international HapMap project, which has catalogued variations at single bases in the genome, termed single nucleotide polymorphisms, or SNPs. Deletions could be identified from apparent inconsistencies in the SNP data, and then tracked through their close association with neighbouring SNPs.

IMMUNOLOGY

Breathe easy

J. Exp. Med. **202**, 1563–1573 (2005)

Treatment with a tiny piece of DNA before exposure to allergens can stop an asthmatic reaction by suppressing the immune system, researchers say.

Edith Hessel of Dynavax Technologies in Berkeley, California, and her colleagues showed that, in mice, the short immunostimulatory DNA molecules act by inhibiting T helper type 2 cells, which normally launch a response against foreign particles entering the body. Immunologists



had previously thought that such treatments acted later, by neutralizing the proteins produced by T helper cells.

The DNA sequences apparently have a second means of blocking allergic airway inflammation: they prevent lung cells from presenting foreign particles to the T helper type 2 cells.

POPULATION DYNAMICS

Deviant sexual practices

Proc. R. Soc. B doi:10.1098/rspb.2005.3370 (2005)

Almost all sexually reproducing species are either hermaphroditic, meaning that individuals have both male and female sex organs, or dioecious, with distinct males and females. But researchers have now found a group of freshwater shrimp species that feature both hermaphrodites and males.

This population make-up, known as androdioecy, is usually thought to occur only as a transition between hermaphroditism and dioecy. But Stephen Weeks of the University of Akron in Ohio and his colleagues, having studied more than 33,000

shrimp from 36 locations around the globe, report that this situation may have persisted stably in the genus *Eulimnadia* for as long as 180 million years.

Phylogenetic analysis of the genus, which features dozens of species (such as *Eulimnadia texana*, pictured left), suggests that androdioecy has survived several speciation events.

MOLECULAR BIOLOGY

Strike while the ion is hot

EMBO J. doi:10.1038/sj.emboj.7600893 (2005)

The mechanism by which pain receptors become more sensitive on stimulation is much debated. A team led by Peter McNaughton of the University of Cambridge, UK, is weighing into the controversy with a new mechanism for how nerve growth factor (NGF) may boost sensitivity to heat by promoting the opening of heat-activated ion channels.

McNaughton and his colleagues show that the binding of NGF to TrkA receptors in the nerve-cell membrane activates the enzyme known as Src tyrosine kinase. This enzyme phosphorylates the ion-channel protein TRPV1, which in turn triggers insertion of the ion channel into the cell surface membrane.

The findings suggest that researchers should revise the model pathway for the activation of TRPV1 channels, which was thought to involve breakdown of the membrane phospholipid PtdIns(4,5)P₂, known as PIP₂.

Correction

In our Research Highlight 'Talking about regeneration' (*Nature* **438**, 534; 2005), credit for the work should have been given to Alejandro Sánchez Alvarado and his team at the University of Utah School of Medicine in Salt Lake City.

JOURNAL CLUB

Peter Crane
The Royal Botanic Gardens,
Kew, London, UK

The director of Kew gardens charts the journey towards mapping the variety of plant life.

I often feel a slight twinge of envy and frustration when confronted with the elegant global analyses of species richness and distributions that are available for birds and mammals.

Such analyses can help to guide

conservation planning. But the sheer number of plant species — more than 20 times the number of birds and mammals combined — make similar studies in my area a daunting prospect.

Increasingly, however, botanists are rising to the challenge. A recent paper from the Nees Institute for Plant Biodiversity at the University of Bonn, Germany, and WWF-US in Washington DC presents the best map available of global plant species richness (G. Kier *et al.* *J. Biogeogr.* **32**, 1107–1116; 2005).

This is a big step forwards —

despite the fact that data for more than half the 867 ecoregions studied was poor, and fewer than a fifth had complete lists of the plant species present.

In reality, I'm pretty certain that knowledge of the distribution of plant species is not quite this bad, but uneven information and the difficulties of drawing it all together present formidable challenges.

What we really need, for science and for conservation, is a working list of all known plant species, linked to fine-grained information

on their global distribution.

Species lists are already available, or are close to completion, for several very large groups of plants, and also for some continents. And for distributions, a vast amount of data already exist, in partly synthesized form, in databases, herbaria and libraries around the world.

The effort required to collate this data is significant, but it is not beyond our reach. Given the pace of change in our environment we should do the job now, while it can still be useful.

NEWS

Big money for cancer genomics

WASHINGTON DC

The US National Institutes of Health (NIH) has launched the pilot phase of its controversial Human Cancer Genome Project, which aims to catalogue the genetic changes associated with cancer.

The project is a potentially huge undertaking that could take 10 years and cost US\$1.5 billion. Its proponents say that tallying up all the genetic mutations in cancer cells may reveal new drug targets.

But opponents argue that cancer biology is too poorly understood to make such a cataloguing approach viable and say that the money would be better spent on basic research into how cancer functions.

The NIH pilot project, announced on 13 December, is itself a major endeavour. Its budget of \$100 million over three years is high, considering that the agency's budget for 2006, which has not yet been finalized, is likely to stay flat or even drop. But Francis Collins, director of the National Human Genome Research Institute (NHGRI), says it would be a mistake to wait. The NHGRI and the National Cancer Institute — both based in Bethesda, Maryland — will split the cost evenly.

"The chance to apply this incredibly powerful engine called genomics to cancer is extremely compelling, and to say, 'Budget times are tough, we're gonna have to wait a while,' would be unacceptable," says Collins.

The pilot plan has five stages. First, the agency will pick two or three types of cancer to study and collect samples of their tumours. Second, it will award money to centres that can do high-throughput analyses, such as gene-expression experiments, on the tumours. Third, the agencies will ask its genome sequencing centres to resequence about 2,000 genes in each tumour. This is part of a larger shift in focus at the NHGRI towards repetitive sequencing of genes associated with disease (see *Nature* **437**, 1233–1234; 2005). The cancer project's scientific advisory board has not yet decided which genes will be resequenced. The fourth and fifth components will be grants for technology development and for bioinformatics systems.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Is this the answer? The NIH is giving strong backing to genomics in cancer research.

The genome project was proposed by the National Cancer Institute's board of advisers earlier this year. But supporters point to earlier success from a similar but smaller effort — the UK Cancer Genome Project at the Wellcome Trust Sanger Institute in Cambridge. In 2002 it turned up a gene called *BRAF* that is mutated in most melanomas and is now a target for drug development (*Nature* **417**, 949–954; 2002). Many cancer researchers and cell biologists, though, say the big-science tactic could suck money away from other grants for years to come. Others caution that the project may not generate useful information unless it includes tests that are not currently planned, such as functional assays that could identify mutations crucial for cancer-cell survival.

Stephen Elledge, a geneticist at Harvard Medical School in Boston, Massachusetts, co-wrote an open letter in October calling on the National Cancer Institute to set up clear stan-

dards by which the pilot project can be judged (S. J. Elledge and G. J. Hannon *Science* **310**, 439–441; 2005). "I would really like to have the sequence of all the cancer genomes, but it may not be that useful and cost-effective," Elledge says. "They need an independent panel of scientists who can evaluate the data and really have the possibility that they will change the way the project is going to go forward."

Collins says that the project's scientific advisory board is looking at this very question and that the agencies will take any recommendations seriously. "We will have very explicit goals," Collins says, adding that finding drug targets should be one of them. And, he says, scientists can apply for money for functional-screening projects as part of the pilot. The pilot will be a crucial test for the institute, and for all genome scientists, as they try to make the case that they can make a difference to medicine. ■

Erika Check

TEK IMAGE/SPL



CLIMATE-CHANGE CONFERENCE BLOG

To get behind the scenes of the meeting, read our diary report from Montreal.

www.nature.com/news

Developing nations offer hope in climate talks

MONTREAL

Delegates from across the world returned home optimistic last weekend after climate-change talks in Montreal, Canada, despite the United States' continuing refusal to commit itself to reducing greenhouse gases.

A key outcome is an agreement by parties to the Kyoto Protocol to discuss deeper commitments to reducing greenhouse-gas emissions, as well as finding creative ways to engage developing countries. Some observers believe that the stage has now been set for long-term cooperation between developing and industrial nations within the treaty.

For instance, Papua New Guinea and several other countries said they want to hold discussions on how to create financial incentives to avoid deforestation in their countries. In addition, China has declared its intent to more than double its use of renewable energy, to 15% of its electrical demand, by 2020. Some at the meeting believe this might lead to non-binding targets within the next commitment period of the Kyoto treaty. China could potentially receive carbon credits if it exceeds that goal.

"These are the kind of innovative things that we now have a negotiating space for countries to put on the table," says Alden Meyer, director of strategy and policy for the Union of Concerned Scientists in Washington DC.

More good news came as talks on the final day, 9 December, lasted into early Saturday morning. Just after midnight, the US delegation finally agreed to consider a dialogue on future strategies. The wording of the text is vague, but the agreement to continue at all was a "tactical



China has targeted renewable energy as an alternative to its conventional power stations.

victory", says Maria Socorro Manguiat, a legal officer at the World Conservation Union.

The United States is not a party to the Kyoto protocol and so is under no obligation to discuss binding targets for emissions cuts. But a proposal made in the first week of talks by the conference president, Canada's environment minister Stéphane Dion, opened up the floor for future discussions under the United Nations Framework Convention on Climate Change, to which the United States is a party.

The Montreal talks ran on parallel tracks: one for countries that are parties to the convention, and one for countries that have ratified the Kyoto protocol. One of the problems with the protocol — particularly for countries such as Australia and the United States that

have not ratified it — is that it mandates binding limits on greenhouse-gas emissions only for industrialized, not developing, nations.

By the end of the talks, parties to the Kyoto treaty had agreed to take part in negotiations towards deeper emissions cuts and other options for stabilizing emissions after 2012, when the Kyoto commitment period ends. They did not agree to a deadline for setting post-2012 commitments, but many observers think there is now a forum in which to engage developing countries such as China, India and Brazil in discussing how to reduce emissions while advancing economic development.

Throughout the conference, the United States insisted that one way forward was through bilateral and multilateral partnerships. One such initiative is the Asia-Pacific Partnership on Clean Development and Climate, which it set up in the summer with Australia and four other countries. This is set to kick off with a ministerial meeting in January. "The partnership will help speed the development of cleaner, more efficient energy systems in some of the world's fastest-growing economies," said Paula Dobriansky, head of the US delegation.

But critics say this is merely a political tool. "At this stage the partnership is simply a skeleton," says Christine Milne, vice-president of the World Conservation Union. The agreement currently has no emissions targets, timelines or funding. "The proof of the pudding will be at the ministerial meeting in Sydney in January," she says.

Amanda Haag

M. HENLEY/IMPACT

Fake pottery buries theory of early start for Christianity

Artefacts recently unearthed in the Netherlands seemed to show that Christianity was openly practised there much earlier than previously thought. But now they have been shown to be fakes.

The archaeologists who dug up the pieces admitted their error last week in the 2005 yearbook of the historical association Numaga. The artefacts, including four pieces of pottery and glass, and a lead platter, were found at different sites in Nijmegen, the oldest city in the Netherlands, between 1995 and 1999. They were briefly displayed

five years ago at Nijmegen's Valkhof Museum.

Most historians think that Christianity was first practised openly after AD 400. But the team of archaeologists, led by Harry van Enkevort, at first dated the items at about AD 200. Other archaeologists expressed doubts about the objects' authenticity at the time of the finds, but it was not until last year that van Enkevort decided to have the finds scientifically analysed.

The tests showed that the inscriptions on the glass and pottery shards were added in the 1990s,

although the shards themselves were from around AD 200. The lead plate was a mere 20 years old.

"You don't expect objects to be fake when you find them embedded in the ground," says van Enkevort. Their appearance had convinced him that they were genuine.

The Dutch scandal is the latest of several recent archaeological scams. Earlier this year staff at the Israel Museum in Jerusalem had to remove from display one of its most prized pieces — an ivory pomegranate that bears an inscription hinting that it had

been used by priests in Solomon's Temple — after finding out that it was fake (see *Nature* 434, 13; 2005). And in 2000 one of the most famous archaeologists in Japan, Fujimura Shinichi, was caught burying artefacts at an archaeological site.

Lothar Bakker, director of the Roman Museum in Augsburg, Germany, and an expert in Christian inscriptions on Roman objects, says that he "wouldn't have trusted the objects" because similar finds have always been later than AD 400. Siëlle Gramser

ON THE RECORD**“It hit an iceberg and it sank. Get over it.”**

Explorer Robert Ballard, who found the *Titanic* in 1985, is unimpressed by the discovery of fresh wreckage from the ship.

“If we are not careful, growing forests could make global warming even worse.”

Climate scientist Ken Caldeira reveals results from computer models that suggest temperate forests absorb sunlight and warm the air.

Sources: Associated Press, Carnegie Institution

SCORECARD**Flying saucers**

Guiyang in China's Guizhou Province has been given US\$20 million by a company in Taiwan to build a UFO research centre. The facility will investigate strange sightings in the area that occurred in 1994.

**Drunk elephants**

A mathematical model has debunked the popular myth that African elephants get tipsy by eating fermented fruit from the marula tree.

**Holy healing**

A Russian scientist has claimed that holy icons speed the recovery of sick mice, but an archbishop has declared his experiments to be sacrilegious.

NUMBER CRUNCH

The Census of Marine Life has unveiled findings from the first five years of its decade-long effort to catalogue all sea life.

3.2 million sightings were this year added to the Ocean Biogeographical Information System (OBIS).

40,000 marine species are now represented by a record in OBIS.

190,000 marine species known to science are not yet represented by an OBIS record.

2.3 million marine species in total are believed to exist.

Source: Census of Marine Life

S. SUBBOTIN/RIA NOVOSTI

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

On the up: Russian space scientists hope they will no longer take a back seat.

Budget boost gets Russia back in the space game

Russia's long-suffering space scientists had reason to celebrate last week as a generous funding increase was approved for the national space agency, giving hope to missions that have long been on hold.

The State Duma — parliament's lower house — approved a budget of 23 billion roubles (US\$800 million) for Roscosmos in 2006, nearly one-third more than the agency received this year. Roscosmos's ten-year budget was set at 305 billion roubles. With oil revenues high, the Duma granted the Kremlin's request for increased spending, and the legislature's upper house is expected to follow suit.

The increase means that Russia may soon return to launching its own space science missions, rather than flying single instruments on European and US spacecraft. The nation's once active planetary programme has been in “miserable shape” for the past decade, following the loss of the Mars-96 orbiter, says Mikhail Marov of the Keldysh Institute of Applied Mathematics in Moscow. Russian space scientists often went unpaid as ideas for missions languished with no hope of reaching the launch pad.

One such mission, called Phobos-Grunt, now seems to be on track to launch in 2009. It will head for the martian moon Phobos, where it will land and collect a soil sample before returning to Earth. The mission has been scaled down — it will use conventional propulsion and launch on a Soyuz rocket, instead of the more expensive Proton — but it should still manage to land 45 kilograms of scientific instrumentation on Phobos.

Spacecraft engineers at the Moscow-based Lavochkin Association are laying plans for an

ambitious mission called Luna-Glob, which would deliver an orbiter and a network of instruments to the Moon for geophysical studies. This mission would probably get funding only after Phobos-Grunt is well under way, says Marov.

Meanwhile, another long-dormant mission, Spectrum, is aiming for launch in 2011 to conduct an all-sky astronomical survey at X-ray wavelengths. Mikhail Pavlinsky of the Space Research Institute in Moscow says the mission is similar “in name only” to a concept called Spectrum-X-Gamma floated in the 1990s, which involved scientists in several European countries and the United States. The list of those involved is now down to Russia, Germany and Britain, with some launch and tracking support from the European Space Agency. Germany hopes to contribute an instrument called eROSITA, originally developed for the International Space Station, to study black holes and other high-energy phenomena.

European space ministers cast the only shadow on this otherwise sunny picture by last week voting not to join Russia in building a new space vehicle. The ministers, who met in Berlin on 5 and 6 December, turned down a request for the €50 million (US\$60 million) needed to join a two-year study of the Clipper space plane proposed by Roscosmos. The ministers did not rule out future cooperation, however. And Russian space officials say they will go ahead with Clipper anyway — although the loss of Europe as a partner would be a major setback.

Tony Reichhardt

Pokémon blocks gene name

A cancer research institute has been threatened with legal action by the US branch of Japanese video-game franchise Pokémon, after one of its researchers borrowed the company's trademark to name an oncogene.

Pier Paolo Pandolfi of the Memorial Sloan-Kettering Cancer Center, New York, first called the new member of the POK family of genes *Pokemon* at a conference in 2001, claiming it was an acronym for POK erythroid myeloid ontogenic. But when Pandolfi and his colleagues described the gene's role in the development of human cancer in *Nature* last January, the discovery attracted headlines such as 'Pokemon's cancer role revealed' (T. Maeda *et al.* *Nature* 433, 278–285; 2005). Message boards and blogs picked up the story, unable to resist using the phrase 'Pokemon causes cancer'.

That led Pokémon USA to exert its legal right to the

trademark, *Nature* has learned. "They threatened to sue us if we did not stop calling the gene *Pokemon*," says Pandolfi, "but the name and the gene have nothing to do with the cartoon." A spokeswoman for Pokémon USA told *Nature* that its image was at risk. "We don't want our image undermined by associating Pokémon with cancer," she said.

This is not the first case of trademark law interfering with a researcher's attempt to name a gene. In 1993, Alfonso Martinez Arias of the University of Cambridge, UK, was told to find an alternative name for his new fly gene *Velcro*, after the Velcro Corporation wrote to the journal that was publishing his paper to say that "such usage invariably dilutes the value of our name and mark".

Perhaps the best-known quirkily titled gene, the fly-development gene *Sonic hedgehog*, has so far escaped legal threats,

despite sharing a name with the spiky electric-blue star owned by that other Japanese video-game giant Sega. Bob Riddle came up with the name in 1993 while working at Harvard University Medical School, but says he doesn't think Sega's image is threatened. "I don't think a development gene harms them," he explains.

Martinez Arias says he has been more careful since his experience with Velcro, but that trademark infringements will continue if geneticists keep looking for catchy names. "They name genes as if they are claiming a new continent," he says. Safe, if boring, systematic names such as those of the Human Genome Nomenclature Committee (HGNC) should be used instead, says Martinez Arias. The Sloan-Kettering centre seems to agree, and is now calling Pandolfi's gene by the HGNC-recognized moniker *Zbtb7*.
Tom Simonite

WARNER BROS. PICTURES

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Creators of the Pokémon world object to the idea of the name being linked to cancer.



STEM-CELL BANK WOULD NEED JUST 150 DEPOSITS

A modest number of cell lines could provide tailored treatments for millions.

www.nature.com/news

ESA



SNAPSHOT Blot on the landscape

Explosions at an oil depot near London on 11 December created one of Europe's biggest ever industrial fires. This image, taken by the Envisat satellite five hours after the initial blast, shows the resulting plume of smoke — it spans more than 140 km. Up to 270 million litres of fuel were held in the 20 tanks that exploded at the Buncefield site.

Atmospheric scientists flew a plane through the smoke on 12 and 13 December to log pollutants present and measure particle sizes. That should help to refine models for how smoke drifts after fires, volcanoes or nuclear blasts.

But the current plume poses little health or environmental hazard. The scale of the blaze is dwarfed by Kuwaiti oil wells set alight in 1991, when around 80 billion litres of oil burned.

Journal grows suspicious of Vioxx data

The integrity of a key study of the painkiller Vioxx has been questioned by the *New England Journal of Medicine*, which published the work in 2000. Vioxx was taken off the market last year because of evidence that it increased patients' risk of having a heart attack or stroke.

In their statement, published online on 8 December, the journal's editors note that three heart attacks in patients taking Vioxx (rofecoxib) were not included in data submitted for a paper about the drug's side effects. The study, sponsored by Vioxx's maker, Merck, compared the drug with the older pain reliever naproxen to see whether it caused fewer gastrointestinal problems. Crucially, the study also documented heart attacks (C. Bombardier *et al.* *N. Engl. J. Med.* **343**, 1520–1528; 2000).

The journal calculates that the risk of heart attack was 4.25 times higher in the Vioxx group. Had the unreported attacks been included, they would have boosted that figure to 5, and quashed the original paper's conclusion that Vioxx adversely affected only those already prone to heart trouble (G. D. Curfman,

S. Morrissey and J. M. Drazen *N. Engl. J. Med.* **353**, 2813–2814; 2005).

Merck says that the published paper accurately described the study's results according to an agreed cut-off date for data analysis. The firm reported the extra events to the US Food and Drug Administration (FDA) in 2000.

In 2001, the editors found out about the events from the FDA, and assumed that they had occurred too late to be included. But when Gregory Curfman, the journal's executive editor, made a deposition in Vioxx-related litigation last month, he discovered that at least two of the authors had known about the heart attacks four months before publication.

A leading epidemiologist says that the authors of the study may not have acted wrongly. "If the outcomes truly occurred after the close of the study, then they don't belong in the study," says Brian Strom of the University of Pennsylvania in Philadelphia. Investigators always establish cut-off dates in their research protocols, but do not always publish details of them in their final papers, says Strom.

In this case, an external committee agreed the cut-off date. But the episode raises questions over whether authors should be required to disclose information on significant events that occur after such a date.

The editors have since discovered that other relevant data were deleted from the manuscript two days before it was submitted. "Taken together, these inaccuracies and deletions call into question the integrity of the data on adverse cardiovascular events in this article," the editors write. They call for a correction.

The paper's lead author, Claire Bombardier of the University of Toronto, Canada, has said that she and the other authors — two of whom are Merck employees — are preparing a response to the editors' statement.

Vioxx was a blockbuster drug for New Jersey-based Merck until it was withdrawn after other work found that it could double someone's risk of having a heart attack. Roughly 7,000 lawsuits have been filed against Merck since the drug's withdrawal.

Meredith Wadman

SPECIAL REPORT

Internet encyclopaedias go head to head

Jimmy Wales' Wikipedia comes close to Britannica in terms of the accuracy of its science entries, a *Nature* investigation finds.



AP PHOTO/M. PROBST

One of the extraordinary stories of the Internet age is that of Wikipedia, a free online encyclopaedia that anyone can edit. This radical and rapidly growing publication, which includes close to 4 million entries, is now a much-used resource. But it is also controversial: if anyone can edit entries, how do users know if Wikipedia is as accurate as established sources such as Encyclopaedia Britannica?

Several recent cases have highlighted the potential problems. One article was revealed as falsely suggesting that a former assistant to US Senator Robert Kennedy may have been involved in his assassination. And podcasting pioneer Adam Curry has been accused of editing the entry on podcasting to remove references to competitors' work. Curry says he merely thought he was making the entry more accurate.

However, an expert-led investigation carried out by *Nature* — the first to use peer review to compare Wikipedia and Britannica's coverage of science — suggests that such high-profile examples are the exception rather than the rule.

The exercise revealed numerous errors in both encyclopaedias, but among 42 entries tested, the difference in accuracy was not

particularly great: the average science entry in Wikipedia contained around four inaccuracies; Britannica, about three.

Considering how Wikipedia articles are written, that result might seem surprising. A solar physicist could, for example, work on the entry on the Sun, but would have the same status as a contributor without an academic background. Disputes about content are usually resolved by discussion among users.

But Jimmy Wales, co-founder of Wikipedia and president of the encyclopaedia's parent organization, the Wikimedia Foundation of St Petersburg, Florida, says the finding shows the potential of Wikipedia. "I'm pleased," he says. "Our goal is to get to Britannica quality, or better."

Wikipedia is growing fast. The encyclopaedia has added 3.7 million articles in 200 languages since it was founded in 2001. The English version has more than 45,000 registered users, and added about 1,500 new articles every day of October 2005. Wikipedia has become the 37th most visited website, according to Alexa, a web ranking service.

But critics have raised concerns about the

site's increasing influence, questioning whether multiple, unpaid editors can match paid professionals for

accuracy. Writing in the online magazine *TCS* last year, former Britannica editor Robert McHenry declared one Wikipedia entry — on US founding father Alexander Hamilton — as "what might be expected of a high-school student". Opening up the editing process to all, regardless of expertise, means that reliability can never be ensured, he concluded.

Yet *Nature's* investigation suggests that Britannica's advantage may not be great, at least when it comes to science entries. In the study, entries were chosen from the websites of Wikipedia and Encyclopaedia Britannica on a broad range of scientific disciplines and sent to a relevant expert for peer review. Each reviewer examined the entry on a single subject from the two encyclopaedias; they were not told which article came from which encyclopaedia. A total of 42 usable reviews were returned out of 50 sent out, and were then examined by *Nature's* news team.

Only eight serious errors, such as misinterpretations of important concepts, were



THE NATURE PODCAST
Listen to Wikipedia founder Jimmy Wales talk about our survey and more at www.nature.com/nature/podcast

D. I. FRANKE/WIKIMEDIA FDN



Kurt Jansson (left), president of Wikimedia Deutschland, displays a list of 10,000 Wikipedia authors; Wikipedia's entry on global warming has been a source of contention for its contributors.



he was the 13th, to more significant inaccuracies. Wikipedia, for example, incorrectly describes how Mendeleev's work relates to that of British chemist John Dalton. "Who wrote this stuff?" asked another reviewer. "Do they bother to check with experts?"

But to improve Wikipedia, Wales is not so much interested in checking articles with experts as getting them to write the articles in the first place.

As well as comparing the two encyclopaedias, *Nature* surveyed more than 1,000 *Nature* authors and found that although more than 70% had heard of Wikipedia and 17% of those consulted it on a weekly basis, less than 10% help to update it. The steady trickle of scientists who have contributed to articles describe the experience as rewarding, if occasionally frustrating (see 'Challenges of being a Wikipedian', below).

Greater involvement by scientists would lead to a "multiplier effect", says Wales. Most entries are edited by enthusiasts, and the addition of a researcher can boost article quality hugely. "Experts can help write specifics in a nuanced way," he says.

Wales also plans to introduce a 'stable' version of each entry. Once an article reaches a specific quality threshold it will be tagged as stable. Further edits will be

made to a separate 'live' version that would replace the stable version when deemed to be a significant improvement. One method for determining that threshold, where users rate article quality, will be trialled early next year. ■

Jim Giles

Additional research by Declan Butler, Jenny Hogan, Michael Hopkin, Mark Peplow and Tom Simonite.

Supplementary information available online at www.nature.com/news/2005/051212/full/438900a.html

detected in the pairs of articles reviewed, four from each encyclopaedia. But reviewers also found many factual errors, omissions or misleading statements: 162 and 123 in Wikipedia and Britannica, respectively.

Editors at Britannica would not discuss the findings, but say their own studies of Wikipedia have uncovered numerous flaws. "We have nothing against Wikipedia," says Tom Panelas, director of corporate communications at the company's headquarters in Chicago. "But it is not the case that errors creep in on an occasional basis or that a couple of articles are poorly written. There are lots of articles in that condition. They need a good editor."

Several *Nature* reviewers agreed with Panelas' point on readability, commenting that the Wikipedia article they reviewed was poorly structured and confusing. This criticism is common among information scientists, who also point to other problems with article quality, such as undue prominence given to controversial scientific theories. But Michael

Twidale, an information scientist at the University of Illinois at Urbana-Champaign, says that Wikipedia's strongest suit is the speed at which it can be updated, a factor not considered by *Nature's* reviewers.

"People will find it shocking to see how many errors there are in Britannica," Twidale adds. "Print encyclopaedias are often set up as the gold standards of information quality against which the failings of faster or cheaper resources can be compared. These findings remind us that we have an 18-carat standard, not a 24-carat one."

The most error-strewn article, that on Dmitry Mendeleev, co-creator of the periodic table, illustrates this. Michael Gordin, a science historian at Princeton University who wrote a 2004 book on Mendeleev, identified 19 errors in Wikipedia and 8 in Britannica. These range from minor mistakes, such as describing Mendeleev as the 14th child in his family when

"Scientists' involvement would lead to a multiplier effect. Experts can help write specifics in a nuanced way."

Challenges of being a Wikipedian

Vaughan Bell, a neuropsychologist at the Institute of Psychiatry in London, UK, has reworked Wikipedia's entry on schizophrenia over the past two years. Around five others regularly contribute to the reworking, most of whom have not revealed whether they have academic backgrounds. Bell says that is not a problem, as disputes are settled through the discussion page linked to the entry, often by citing academic articles. "It's about the quality of what you do, not who you are," he explains.

While admitting it can be difficult

settling arguments, Bell says he often learns something by doing so. One user posted a section on schizophrenia and violence that Bell considered little more than a "rant" about the need to lock up people with the illness. "But editing it did stimulate me to look up literature on schizophrenia and violence," he says. "Even people who are a pain in the arse can stimulate new thinking."

Others, particularly those who contribute to politically sensitive entries, have found the editing process more fraught. William

Connolley, a climate researcher at the British Antarctic Survey in Cambridge, has fought for two years with climate-change sceptics over the entry on global warming. When Connolley was insulted by one of the sceptics and the editing became a 'revert war' — where editors repeatedly undo each others' changes — the matter was referred to the encyclopaedia's administrators.

Two of Connolley's opponents were banned from editing any climate article for six months, but it was a bumpy process. The

Wikipedia editors who oversaw the case took three months to reach a decision. They also punished Connolley for repeatedly changing the sceptics' edits, placing him on a six-month parole during which he is limited to one revert a day. Users who support Connolley have contested the decision.

"It takes a long time to deal with troublemakers," admits Jimmy Wales, the encyclopaedia's co-founder. "Connolley has done such amazing work and has had to deal with a fair amount of nonsense."

J.G.

Hayabusa didn't grab asteroid sample after all

Just 11 days after celebrating the world's first collection of a piece of asteroid, the Japan Aerospace Exploration Agency (JAXA) has had to take back its words. On 7 December, JAXA announced that the Hayabusa spacecraft apparently failed, after all, to pick up a sample from the asteroid Itokawa.

Newly retrieved data suggest that the probe may not have fired the two pellets that were supposed to pulverize Itokawa's surface and allow fragments to be collected. Jun'ichiro Kawaguchi, the project manager, said that a safety system might have accidentally switched on and prevented the probe from firing the pellets.

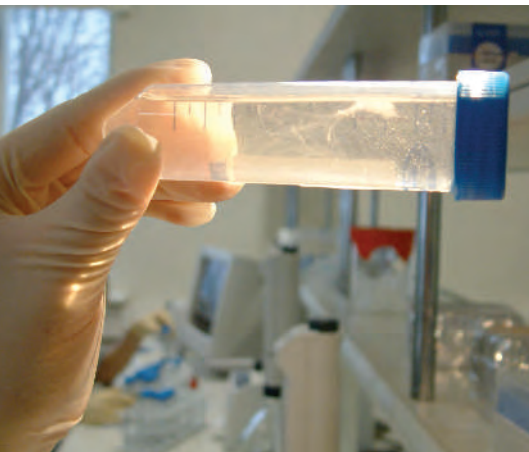
Hayabusa is also having serious engine trouble that affects its power and attitude control. Nevertheless, JAXA still hopes to get the troubled probe back to Earth in 2007.

Cash infusion saves Estonian gene bank

The Estonian government has decided to bail out the tiny eastern European country's ambitious gene-bank project.

The Estonian Genome Project, founded in 2001, aims eventually to collect and bank blood samples from most of its 1.4 million citizens, together with medical and lifestyle data. When venture-capital funding for the scheme dried up last year, after only 10,000 blood donors had joined up, the project was left treading water.

The government now plans to inject €8 million (US\$9.5 million) into the project over the next four years, enough to raise the number of donors to 100,000. This size would make the gene bank a serious player in population genetics, where scientists study the complex interplay of genes and environment underlying common diseases such as high blood pressure and depression.



Give a little bit of yourself: the government hopes every Estonian will eventually be in its gene bank.

Norwegian children's not-so-mad scientists

Goodbye, Dr Frankenstein. Kids these days have a more multifaceted image of scientists, according to some of their pictures.

A decade ago, when the Research Council of Norway's youth club held an art competition for children aged 5 to 15, most of the pictures of scientists were of the stereotypical mad variety, complete with dishevelled hair and test tubes. And they were all men.

In this year's competition, though, scientists were drawn excavating dinosaur bones, in outer space and studying whales underwater. Of the hundreds of drawings submitted, 30% were of female scientists.



The council's Marianne Løken says she hopes the drawings reflect more realistic views of scientists, but cautions that the study is just an informal glimpse of children's attitudes.

Australian law could force nuclear dump on territory

Australia has moved closer to selecting a permanent site for the nation's first nuclear waste dump. On 8 December, the Australian government passed legislation allowing the facility to be built at one of three sites in the Northern Territory.

The decision has angered local communities and indigenous people's groups, and follows years of wrangling between federal and state governments over where to put the site. State political pressure killed an earlier proposal to house the facility in South Australia. The new federal legislation will overrule any opposition from the Northern Territory government.

The site will serve as a repository for nuclear waste from agencies such as the Australian Nuclear Science and Technology Organisation, which runs the country's only nuclear reactor near Sydney. Australia's nuclear waste is currently stored in universities, hospitals and research facilities across the country.

At last India joins the nuclear-fusion club

India has joined the multi-billion-dollar nuclear-fusion experiment ITER, becoming the seventh member of the consortium.

ITER's participants — the United States, the European Union, Russia, Japan, South Korea and China — made the decision to admit India on 6 December during a meeting in South Korea. The offer comes nearly two years after India's request to join.

ITER aims to demonstrate the feasibility of controlled self-sustaining nuclear fusion

as an energy source. The main ITER facility will be built in Cadarache, France, and all ITER partners will participate in its construction, development and research.

Prizewinning homeopathy research is withdrawn

A pharmacologist at the University of Leipzig in Germany has admitted errors in a study claiming that a form of homeopathy worked. The errors, described in a university statement, include missing control experiments and an incomplete statistical analysis.

Karen Nieber was the lead author on a study that claimed that a homeopathic dilution of the drug atropine could relax cramps of gut muscles in rats (K. Nieber *et al. Biologische Medizin* 32–37; February 2004). Before the work was published, it won a homeopathy research prize.

Nieber says that she has since retracted the paper and plans to pay back her share of the €10,000 (US\$11,900) prize. The university began investigating the work after other researchers pointed out suspected flaws.

Correction

Of the 11 cell lines described in Hwang *et al.* (W. S. Hwang *et al. Science* 308, 1777–1783; 2005) mentioned in our News story "TV tests call into question cloner's stem-cell success" (*Nature* 438, 718; 2005), it was originally claimed that seven could divide into different cell types, a figure subsequently revised down to three in a formal correction. And the text specifying the contribution of Gerald Schatten and his colleagues at the University of Pittsburgh was amended during routine copy-editing between the paper being posted online in May and its appearance in print in June.

THE CHAOS TO COME

Natural disasters have wreaked havoc on the planet in the past twelve months, and some say that things will only get worse. **Quirin Schiermeier** assesses the world's growing vulnerability to catastrophe.



CREDIT

In sheer numbers, the death toll of floods, earthquakes, tsunamis, volcanoes and cyclones is small — 80,000 in an average year. Even the recent spate of catastrophes ranks surprisingly low among the scourges of humanity. With the Indian Ocean tsunami and the Kashmir earthquake, disasters in the past 12 months claimed more than 400,000 lives — the highest toll since 1970. But more than three times that number are estimated to have died on Earth's roads in the same period; more than twenty times as many died of avoidable childhood diseases.

To dwell on the averages, however, is to miss the point. Catastrophes are not average; they are the great exceptions. Most of us have seen road-traffic accidents, but few of us have witnessed a natural disaster. They are ruptures with the everyday that change cities, countries and sometimes whole regions forever. They can even change patterns of thought. The Lisbon earthquake of 1755 and its associated tsunami, which struck while the city's churches were celebrating All Saints' Day, shook the faith of millions and altered that century's intellectual landscape irrevocably.

It is the fact that they are rare and exceptional that makes it very hard to plan for natural disasters. Problems that crop up less than once a generation — even, in some cases, less than once a millennium — are easily overlooked. Experts can anticipate some natural disasters, but their predictions and assessments often count for little compared with the pressures of

population and development, which increase the number of people in harm's way. As more people crowd the planet, particularly in vulnerable coastal areas, they risk disasters of unprecedented proportions. Hence the paradox that the twentieth century, benefiting from the most advanced technology in human history, saw more deaths due to natural disasters than any previous century had. There is even the chilling possibility that the current century will beat the previous one's tally of 3.5 million deaths.

The science behind predicting disasters is uncertain. But with the best geological and meteorological knowledge in hand, along with models that assess the vulnerability of different populations (see 'Insuring for disaster', overleaf), experts are starting to quantify just how grim the future will be. Steve Sparks, a volcanologist at the University of Bristol, UK, has looked at analyses of current trends by the German reinsurance firm Munich Re. "The world can now expect three to five major events per year, which each kill more than 50,000 people," he says.

The mechanisms behind these disasters are not new; the basic factors that underlie volcanic eruptions, earthquakes, cyclones and floods are well understood, and so is the geographic distribution of risk. But there are still three mysteries: where and when a disaster will strike; what unforeseen consequences there might be; and how to get policy-makers and the public at large to take the problems, and the uncertainties, seriously.



TSUNAMIS

Past More than 60,000 people were killed in Lisbon in 1755 by an earthquake and the resulting tsunami (pictured). About 300,000 have been reported dead or missing following the great Indian Ocean tsunami of December 2004. Tsunamis can be triggered by a range of disturbances, including earthquakes, underwater landslides and asteroid impacts.

Future Low-lying coastlines in Sumatra, China, Peru and the eastern Mediterranean are considered at greatest risk. Early-warning systems, public education, and land-use regulation offer the best chance of saving lives.

AP/EMPHICS

BETTMANN/CORBIS

Public preparedness rarely keeps pace with scientific knowledge. "The science," says Sparks, who recently organized a meeting on natural disasters at the Royal Society in London, "is not getting through to policy-makers, planners and populations." US meteorologists had often warned that their worst-case hurricane scenario would involve a storm breaking the protective levees around low-lying New Orleans. And that's exactly what happened on 29 August with Katrina, which left 1,300 dead — the largest death toll from a natural disaster in the United States since the 1928 Okeechobee Hurricane. Geologists had long warned of the possibilities of large earthquakes along the Himalayan front, like the one that struck Kashmir in October, and in the Sumatran trench off Indonesia, where the tsunami was born last December.

The very enormity of the Indian Ocean tsunami is, in fact, a good example of unforeseen consequences. The seismic risk in the area was known, but the scope of the resulting tsunami, which affected a dozen countries, was totally unexpected. Unlike in Hawaii and Japan — where tsunamis are a relatively frequent phenomenon, and where education projects and early-warning systems exist — many people in southeast Asia, India and eastern Africa had never even heard of such waves. The death toll was particularly high in the 'shadow zone' near the Equator, where tropical cyclones rarely occur and people have lived safely by the sea for centuries.

Risk-management experts predict that a tsunami of similar scale is likely to happen only once every 500 years. Nevertheless, nations around the Indian Ocean are funding an early-warning system similar to the Pacific Tsunami



HURRICANES

Past The Great Hurricane of 1780 killed 22,000 in Martinique, St Eustatius and Barbados. In 1900, an unnamed storm wiped out the Texas island of Galveston (pictured) establishing Houston as the state's main port.

Future Experts point to several scenarios that could be worse than August's Katrina, including a category-5 hurricane racing up the US east coast and into Manhattan. Low-lying parts of eastern India and Bangladesh are at high risk of large fatalities from flooding and storm surges. Population growth along coasts increases vulnerability.

Warning Center in Hawaii — the first set of buoys was installed last month. Critics charge that the technology could rust at the bottom of the ocean long before the next big wave strikes¹, but to some extent this misses the point. Many holders of insurance never need to make a claim, but that doesn't mean it was silly to buy a policy.

History lessons

Letting the past be a guide to the future is not entirely satisfactory. Hemant Shah is the president of Risk Management Solutions (RMS), a company based in Newark, California, that constructs risk models for the insurance industry. He echoes the US defense secretary, Donald Rumsfeld: "The challenge for the industry really is to think about the unknown unknowns." More thinking along those lines might have led to a greater appreciation of the tsunami risk in the Indian Ocean. But analysis of the past does at least provide data for imagining future scenarios — experts can envisage the consequences of a repeat of the Lisbon earthquake, for example.

A team led by Shinji Toda, of the Geologic Survey of Japan in Ibaraki, and Ross Stein, of the US Geological Survey in Menlo Park, California, has been re-evaluating seismic risks to Tokyo in light of past earthquakes. In a report submitted to the Japanese government, the researchers found that a replay of the 1855, magnitude-7.3 Tokyo earthquake that killed 7,000 people could devastate the city and send shocks through the world's financial markets. The problem is that no one understands the 1855 earthquake well enough to predict accurately when it

Insuring for disaster

The impact of natural disasters on the insurance industry often bears little relation to death tolls. A pair of hurricanes from 2005 — Katrina and Wilma — may together cost insurance companies a record US\$40 billion. And European winter storms with almost no fatalities can still cost billions.

Such bills make modelling risk vital for insurers. Models help companies adjust their policy prices to their 'risk appetite', decide how much reinsurance to buy and how much additional capital to hold.

In 1995, Lloyd's of London began asking its 66 underwriting syndicates to report estimated losses from hypothetical disasters based on historical examples. Scenarios included major earthquakes in California, the US Midwest and Japan, and windstorms in Florida and Europe. The resulting 'realistic disaster scenarios' help Lloyd's to assess the insurance market's

vulnerability to big losses.

With few major natural disasters in the 1970s and 1980s, the insurance industry was ill-prepared for 1992's Hurricane Andrew and 1994's Northridge earthquake. Together, these cost insurers

can calculate the frequency of a certain size of event (see table) and tell an insurance company with a given portfolio how big a loss it can expect.

The models usually consist of three modules. A high-resolution

calculates an insurance company's effective losses given such damage, taking into consideration its portfolio and the location of its policies.

The models vary widely in robustness. They work well for calculating losses from hurricanes and earthquakes in the United States. But in places where there is little information from historical disasters, from Australia to Morocco, risk modelling is difficult and uncertain.

And for some of the most apocalyptic scenarios — a meteorite impact or a mega-tsunami in the Atlantic Ocean — modelling doesn't work at all. "It is a real concern for the insurance industry that no models exist for some of the worst things that could happen," says Iwan Stalder, head of catastrophe perils at Zurich Financial Services in Switzerland. "But we have to accept that some risks are simply unquantifiable for underwriting."

Q.S.

Frequency of event	Size of event (measured in fatalities)		
	Earthquake	Flood	Volcano
1 in 100 years	472,000	98,000	34,000
1 in 500 years	1,052,000	520,000	74,000
1 in 1,000 years	1,446,000	1,061,000	97,000

Data provided by Risk Management Solutions

\$30 billion; at least nine companies were rendered insolvent. Other crises, such as the 1995 Kobe earthquake in Japan and the 2001 terrorist attacks on the United States, have also had an effect — over the past decade, the cost of reinsurance has grown 40-fold.

To help insurance companies cope, risk-management specialists draw up models of disasters. They

hazard module predicts the physical effects that might be expected at a given point from sources such as historical records, windspeed maps or seismic data. A vulnerability module then predicts the damage that an event, depending on its strength and duration, would do to local buildings and infrastructure. Finally, a financial module



AFP/GETTY IMAGES

might happen again, and thus to quantify the risk. "There's an enormous amount of uncertainty," says Stein.

Tokyo's peculiar vulnerability to earthquakes, which arises from its position at the junction of three tectonic plates, is a chronic worry for the global financial system. But in casualty terms, there are places that concern seismologists more. The worst earthquake of the twentieth century, and possibly the worst in history, was China's Tangshan earthquake in 1976, which may have killed more than half a million people. Some seismologists think that the ever-growing cities in earthquake belts across Asia, many filled with unstable buildings, mean that a million-fatality earthquake is possible in the near future; Sparks mentions Istanbul, Tehran and Sumatra's Padang as likely sites. Tehran, a city of 12 million with what one expert at RMS calls "some of the scariest construction known to man", is a particular worry: just as Tokyo has the potential for an earthquake with disproportionate financial impact, one in Tehran would have a huge geopolitical impact.

Floods of fire

Volcanic eruptions are also bound to do more damage in the future than in the past, simply because more people live in their path. Here some degree of warning is more likely than it is with earthquakes, but it cannot be relied on. And although cities such as Naples, at the foot of Italy's Mount Vesuvius, face the most obvious dangers, other situations could result in more deaths. Robert Muir-Wood, chief researcher at RMS, notes that a major tragedy could happen if an eruption reached a major city just outside an immediate evacuation zone, to which people had fled for shelter.

An additional concern is that we have no historical experience of what a really large

Shattered: the destruction caused by the Indian Ocean tsunami was largely unforeseen.

volcanic eruption might look like. The twentieth century's most lethal eruption, in Martinique in 1902, was far smaller than the giants of the nineteenth century. The 1883 eruption of Krakatoa in Indonesia obliterated the island and sent plumes of volcanic ash drifting around the globe for years, and the Tambora eruption of 1815 affected climate all over the world. Even mighty Krakatoa was at least an order of magnitude smaller than the Santorini eruption, which effectively wiped out the Minoan civilization in the eastern Mediterranean around 1400 BC.

Hurricanes, by contrast, are more tractable. There are limits to their size and destructive power, they have seasons, and their tracks can be forecast with some degree of accuracy. They are also frequent enough for people to take seriously, especially if given the right sort of warnings and information. In the United States, the National Hurricane Center in Miami, Florida, is responsible for issuing storm warnings when Atlantic hurricanes threaten the coast. It now produces maps, available on the web and through television broadcasts, that forecast storm paths with an accuracy that has improved by 50% in the past 30 years.

Most coastal residents pay attention only when a hurricane is threatening their particular part of the coast, so this information has to be targeted to the right place, at the right time and in the right way. "Meteorology means nothing to 99% of my viewers," says Steve Lyons, a hurricane expert and anchorman of the popular US Weather Channel. "We need to translate the meteorology into impact. It's like telling people in a town very concretely: 'It's gonna blow your roof off, so you better run.'"

But it is not all about listening to the media. Teaching people how to notice warning signals from nature — such as shaking ground or a



VOLCANOES

Past The past two centuries have seen more than a dozen volcano disasters that caused at least 500 deaths. One such was Mt Pelee's eruption in 1902 (pictured). Affected regions are mostly in the developing world, including Indonesia, Colombia, Mexico, Guatemala, Martinique, Papua New Guinea, Cameroon and Congo.

Future Were the 1815 eruption of Indonesia's Tambora to happen today, 15,000 people might be killed. Volcanologists are struggling to develop timely warnings for mountains that have lain dormant for many years and could erupt at a moment's notice.

POPPERFOTO.COM

AP/EMPHICS



The death toll of the Kashmir earthquake is still rising.

receding ocean — can save lives in all sorts of natural disasters. Other relatively simple measures include planning escape routes and evacuations, and reconsidering land use in regions threatened by floods, storms, landslides and avalanches. A 2004 World Bank report on the cost of natural disasters in the 1990s estimates that \$40 billion invested in risk reduction and preparation could have cut the decade's final bill in half, from \$535 billion to \$255 billion. It also estimates that, over the past four decades, \$3.15 billion invested in flood control by China has averted \$12 billion in losses.

Such mitigation strategies have worked to save lives most dramatically during floods, says Muir-Wood. In China, for instance, flood casualties have dropped throughout the century, in part because of investment in protection systems and evacuation plans. In the 1930s and 1940s, 4.4 million people died from flooding in China; in the 1950s and 1960s, that number dropped to 2 million, and by the 1970s and 1980s it was 14,000.

Still, the Indian Ocean tsunami and Katrina have highlighted the failure of society as a whole to take preventative action, says Maxx



KEYSTONE/GETTY IMAGES

FLOODS

Past Seven of the ten deadliest flood disasters in the twentieth century have occurred in China. More than 6 million people died from drowning, starvation and disease during the three biggest floods in 1931, 1939 and 1959.

Future China is building its controversial Three Gorges Dam, to be completed in 2009, in part to reduce flood risk. Central America, Bangladesh, Taiwan and elsewhere remain at risk from heavy rains. Dams and other barriers, as well as re-naturalization of river courses, may help. Low-lying countries such as the Netherlands are accustomed to inundations such as that pictured above, but climate change may increase flood risk.

Dilley, a geographer at the Disaster Reduction Unit of the United Nations Development Programme (UNDP) in Geneva. From international organizations to local decision-makers, those in charge are realizing that they need to know how natural events become tragedies. "A lot of things are coming to a head," says Dilley. "There is an international push to focus more on risk identification and management, as opposed to post-disaster emergency action."

Several organizations, including the UNDP, the Red Cross and the International Strategy for Disaster Reduction, have recently published reports outlining improved risk-management strategies²⁻⁴. They emphasize putting more resources into preparedness — restricting the expansion of cities in earthquake-prone regions, for example. Science, says Dilley, should be the basis of all these activities. But the fragmented research community is not positioned to provide the right input, he adds: "What you need are planner-friendly common views, as opposed to highly specialized scientific papers."

To address the issue, a group of experts led by David King, Britain's chief science adviser, has suggested setting up an International Science Panel for Natural Hazard Assessment. Such a panel would function like the Intergovernmental Panel on Climate Change, sorting through a flood of scientific information and creating a generally accessible summary of the latest findings. And the UNDP and World Bank are working with Columbia University's Earth Institute in New York on a project to identify high risks, and give governments and local planners the data to recognize them. "This is the kind of information," says Sparks, "that, had it been available ten years ago, would have allowed governments in southeast Asia to take precautions against the tsunami."

Quirin Schiermeier is Nature's German correspondent.

BETTMANN/CORBIS

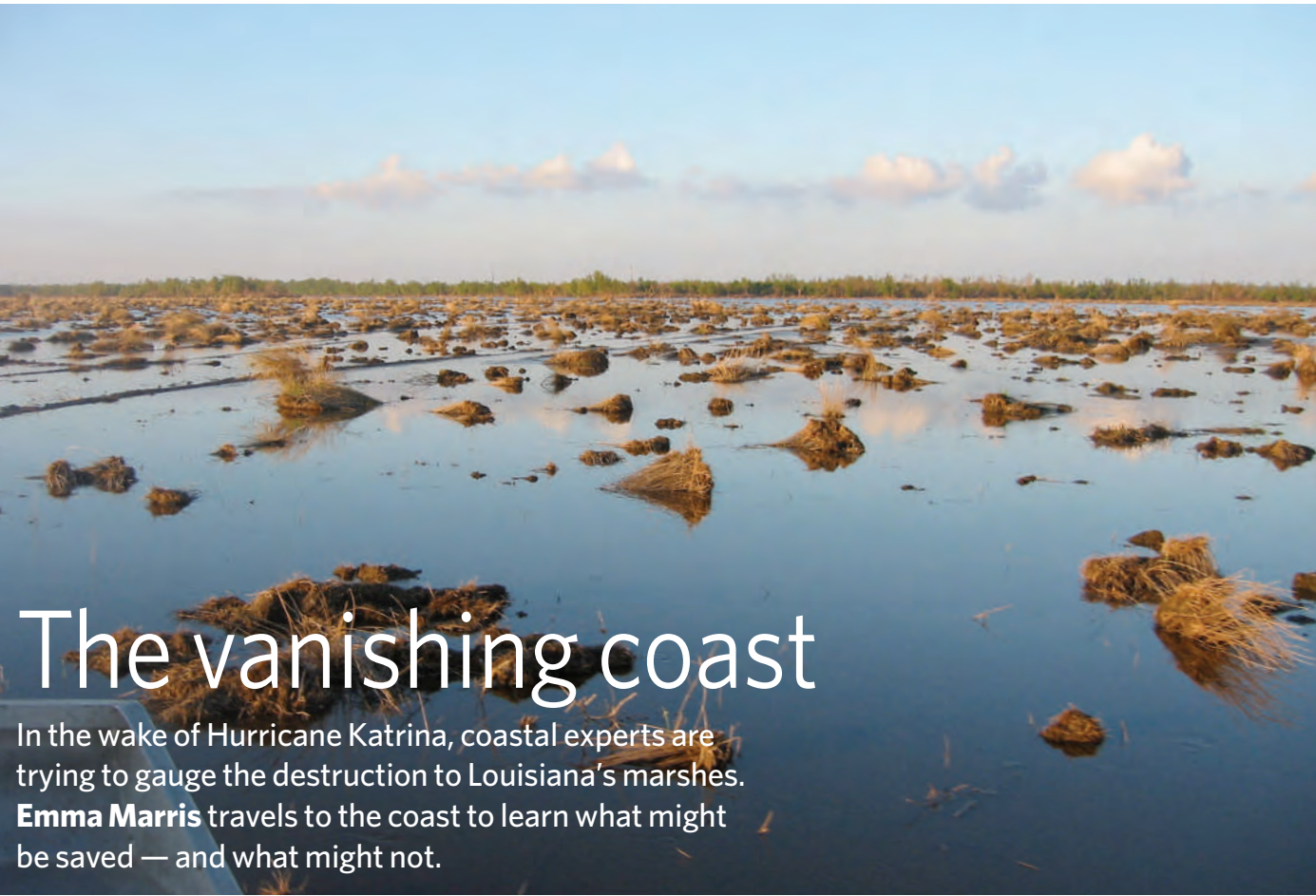


EARTHQUAKES

Past The devastation of San Francisco in 1906 (pictured) is legendary, but the deadliest earthquake known happened in China: at least half a million died in the Tangshan disaster of 1976.

Future Specialists point to China, the Himalayan front or the city of Tehran as possible locations for the first earthquake to cause a million casualties. Events in Tokyo, California and the Mississippi Valley are likely to cause the most economic damage. Improving construction standards and emergency plans could significantly reduce fatalities.

1. Alversen, K. *Nature* **434**, 19–20 (2005).
2. *Reducing Disaster Risk: A Challenge for Development* (United Nations Development Programme, 2004).
3. *World Disasters Report 2005* (International Federation of Red Cross and Red Crescent Societies, 2005).
4. *Living with risk: A global review of disaster reduction initiatives* (International Strategy for Disaster Reduction, 2004).



The vanishing coast

In the wake of Hurricane Katrina, coastal experts are trying to gauge the destruction to Louisiana's marshes.

Emma Marris travels to the coast to learn what might be saved — and what might not.

Getting stuck was the highlight of the day. In Little Vermilion Bay, along Louisiana's fast-disappearing coast, researchers cherish every bit of land they can find. So when a boatload of scientists recently got stuck in chocolate-brown mud, they celebrated even as they struggled. "We got the boat through here last time," says Richard Raynie, a coastal-resources scientist at the Louisiana Department of Natural Resources (DNR), satisfied with the change.

The Louisiana coast, a lacy delta plain known for its seafood, oil, alligators and hurricanes, is unravelling. Generations of engineering projects, designed to make the soggy floodplain habitable, have disrupted the processes that held the wetlands together.

Water that once deposited sediments onto the Mississippi River Delta is now hustled into the Gulf of Mexico by levees and floodwalls. Canals let the sea into coastal marshes, bringing salinization and erosion.

Hurricanes Katrina and Rita, which hit opposite ends of Louisiana this August, underscored the marsh's fragility. The storms exposed the choices that must be made about land used for habitat, real estate, hunting, fishing, seafood and oil. Now, with the momentary but fading attention of the nation, scientists who have long worked to beat back the sea are being asked to solve every problem at once: Can we save everything? What do we

save first? What are the most effective methods? Where should the money come from?

The resources of the marsh, as well as the infrastructure located there, are worth a lot in Louisiana, and the state has long fought to keep its land from dropping into the sea. But in the 1990s, Louisiana lost just over 60 square kilometres of coastal wetland each year — an area the size of a football field every half-hour. Katrina and Rita washed away another 260 square kilometres, according to early estimates by the US Geological Survey (USGS).

New bayou

Engineers have built a series of projects to try to hold on to the land. Among them is the Caernarvon water diversion, 25 kilometres south of New Orleans, where river water pours through a levee onto wetlands.

It is a pale imitation of the flooding that once fed this area with sediment and freshwater. The diversion often flows at far less than its maximum capacity of 230 cubic metres per second — and since Katrina, it has not flowed at all because of a lack of electricity and other problems. Even under ideal conditions, estimates suggest, it will save only 65 square kilometres of land over 50 years.

Even so, there are signs that Caernarvon is bringing freshwater marsh back to the region. As researchers' airboats glide across the water, dozens of alligators wake up. Freshwater plants,

such as alligator weed and rattlebox, thrive. Roseate spoonbills mix with other birds.

But since Katrina blitzed through this area, open water and soggy land have swapped places. The wind picked up acres of marsh and flung it about in tiny handfuls called marsh-balls. Researchers are debating how many — if any — of these balls will take root in open water, perhaps creating new marshland.

The Caernarvon diversion opened in 1991, a year after Louisiana senator John Breaux got funding for coastal restoration. Now, a system of federal and state trust funds pays for such projects; even so, they restore only a small fraction of the lost wetlands.

The Breaux money funded a series of restoration projects without an overarching vision. "People began to ask: 'Do we have a big plan?'" says Gerald Duszynski, head of the DNR's office of coastal restoration and management. The result was a multi-agency plan called Coast 2050, stuffed with goals — protecting shoreline here, building a delta there, and diverting water all along the Mississippi.

Not surprisingly, the federal government balked at the \$14-billion price tag. A short-term, \$1.9-billion plan was drawn up. But the projects in that plan are not the integrated approach that the coast's problems demand, according to an evaluation released in November by the US National Research Council.

Meanwhile, some of the projects made

headway. In Little Vermilion Bay, near the Texas border, a \$900,000 project built 23 long, thin islands, known as terraces, to slow coastal erosion. Over the past few years, a whole cast of typical marsh plants has arrived, perhaps just in time: Hurricane Rita hit the area, but the terraces were thickly vegetated and so suffered only minimal erosion.

On a hot November day, DNR scientists David Castellanos and Dona Weifenbach work their way across the islands, recording every plant in metre-long segments. "*Alterniflora*, bull tongue, *Phragmites*, deer pea, alligator weed," Castellanos calls out. "We encourage everything that will hold its roots down," says Weifenbach.

The terraces work in many ways. "We protect the shorelines, but we also gain acreage," says Castellanos. By catching sediment, the terraces create additional land. A little less than half a metre was laid down between 1999 and 2003. Now — as Raynie and his colleagues discovered — there is enough to beach a boat.

Different strokes

Few dispute that Louisiana should fight to keep its land; hunters and fishermen value the marsh as much as the environmentalists. But there is disagreement over what kind of strategies should be used. Put simplistically, engineers prefer hard structures such as levees and rock walls, biologists prefer freshwater diversions such as Caernarvon, and geologists prefer soft earthworks such as terraces.

Each approach has its drawbacks. "For freshwater diversion really to work, it takes decades," says Shea Penland, a geologist at the University of New Orleans. Robert Twilley, a systems ecologist at Louisiana State University in Baton Rouge, agrees that diversions take time, but argues that they are worth doing because they run cheaply and improve the marsh. Expensive earth-moving exercises are "short-term triage," he says — something to keep the problem from getting worse.

One problem is where to get the earth from. "You can't just rearrange the resources — you need to bring in new resources," says Twilley. Some experts want to dredge up sediment from the sea floor and add it to the coast. In particular, a number of restoration experts have their eye on a huge vein of sand called Ship Shoal, 20 kilometres offshore.

These include Mark Kulp, also of the University of New Orleans, who is in a floatplane over Plaquemines Parish south of the city. As one goes south, the strips of land get narrower and the wreckage worse. Ten weeks after Katrina, an ocean barge lies on its side on a levee. Houses are a few heaps of lumber on a slab, and the ground twinkles with broken glass.

In Kulp's ideal world, this scenario would never happen again. The communities on the lower part of Plaquemines Parish would be moved to higher ground. "We need to be willing to cut our losses," he says. He calls this idea managed retreat, prioritizing which bits of



Swamped: scientists from Louisiana's Department of Natural Resources survey Little Vermilion Bay.

land are worth the money and energy to save. The National Research Council study noted that some people may have to leave vulnerable areas. The ideal time to move people, Kulp suggests, would be now, as so many were uprooted by this year's storms.

It's not clear whether the wetlands' sudden fame will improve the fortunes of the restorers. In June, Louisiana politicians scored nearly

\$90 million annually over four years for coastal restoration, in addition to the trust fund set up by the Breaux act. Legislators included \$1.4 billion for coastal restoration in an early version of a budget bill. It remains to be seen if it will be part of the final version, which lawmakers hope to get out before the end of the year.

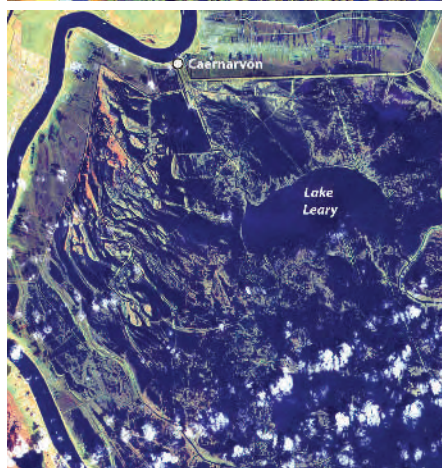
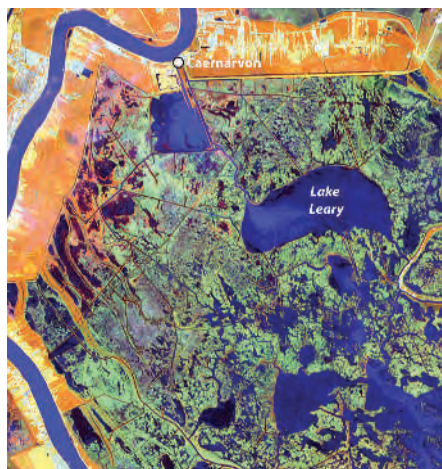
In rehab

Since Katrina, the state legislature has voted to combine hurricane protection and coastal restoration under one authority and, potentially, one trust fund. This makes Duszynski nervous. "We're worried that the pressure is going to be so heavy to build levees that the restoration will be left behind," he says. On the other hand, any windfalls for levee-building may possibly get redirected to restoration work. But money has to come soon, he warns. "The problem is that as time ticks on, the pressure is less to do anything," he says, "and then we are back to begging for restoration money."

At any rate, restoration is the wrong word for what these scientists are aiming for. None of them expect the levees and floodwalls to go. There is no one agreed map of what the coast should look like in the future. "You can restore, you can rehab, you can preserve," says Chuck Villarrubia of the DNR. "Really what you want to do is set the habitats on a trajectory that is sustainable."

Castellanos puts it another way: "The natural system has been altered, and you can't go back. The bottom line is that every time you increase that acreage, you get benefits — forage for fish, shoreline protection, and even aesthetic values for people." For every researcher who secretly wishes people would leave the marsh to the alligators, there is another who wants to use the tools of the restorer — rocks, sand, freshwater and mud — to engineer the coast they want. Coastal restoration, it is clear, is mostly coastal design.

Emma Marris reports for *Nature* from Washington DC.



Before and after: satellite images show marshland lost in Breton Sound, Louisiana, one of the areas hit hardest by Hurricane Katrina.

E. MARRIS

USGS NWRC

ROOTS OF RECOVERY

Replanting coastal forests destroyed by last December's tsunami sounds like a good idea — especially if they protect against future storms. But such plans need nurturing if they are to succeed, **Erika Check** discovers.

On a cloudy, humid day last month, 250 people gathered near the beach in the small Indonesian village of Kajhu. Local and international officials gave speeches, and some of the villagers performed a traditional dance ceremony. Then, everyone carried shovels and buckets along a gravel road past a flattened area of the beach littered with broken tree stumps. The buckets held tiny mangrove and *Casuarina* pine seedlings, which the villagers tucked into shallow trenches they dug at the water's edge.

Kajhu village is in the province of Aceh on Sumatra — the area most brutally devastated by last year's tsunami. Of the approximately 300,000 killed on 26 December 2004, more than 130,000 were from Aceh. The Kajhu ceremony was a small act of hope in the wake of tragedy. The replanting marked the launch of the Green Coast Programme, a plan drawn up by non-profit groups, international donors and the Indonesian government. The project is one of many across southeast Asia that aims to replace mangrove forests destroyed by the tsunami and by human industry.

Mangroves have been disappearing from southeast Asian coastlines for decades, replaced by vast shrimp farms and tourist resorts. In the five countries hit hardest by the tsunami, development eliminated 1.5 million hectares of mangroves between 1980 and 2000 — 26% of the region's mangrove cover¹. But the devastation wrought by the tsunami has inspired governments to try to restore these ecosystems, which environmental scientists and economists have long said are natural defences against storm damage.

Mangrove restoration is notoriously difficult, however, and these replanting projects face huge challenges. Past restoration projects have planted the wrong species in the wrong places, scientists say. And the projects have tended to ignore, alienate or exploit the people living closest to the mangrove forests. "Local communities must have more of a say in the control, use and protection of mangroves,"



Green shield: workers near Banda Aceh plant mangroves in the hope of protecting their coastline.

says Ed Barbier, an environmental economist at the University of Wyoming in Laramie. Otherwise, future restoration projects will suffer the same fate as past ones — and fail to have any lasting results.

Boom and bust

The problem, says Alfredo Quarto, director of the Mangrove Action Project, Port Angeles, Washington, is that "mangrove areas are remote, usually public lands, available to lease by corrupt officials". Poor fishers and farmers rarely have any land rights and cannot prevent mangroves being cleared for shrimp ponds. "The people who enforce the laws don't live in these areas and can be convinced by someone with money to turn their backs on the destruction," Quarto says. When the farms collapse,

due to disease or contamination, a wealthy owner can move on to another stretch of virgin coast, leaving a useless waste site behind².

Shrimp farms create temporary jobs, but the boom-and-bust cycle hurts villages more than it helps. Healthy mangroves are nurseries for fish, crabs and other creatures that mature in the ocean. So eliminating the mangroves cuts off a source of food and income.

Mangrove deforestation compounded the effects of the tsunami. After the wave passed, villagers reported that areas with more mangrove cover suffered less damage. These anecdotal reports have been bolstered by analysis of satellite images taken before and after the tsunami. This October, international researchers coordinated by the Nordic Agency for Development and Ecology in Copenhagen,

B. BAKKARA/AP

Denmark, reported that areas with mangrove or tree cover were significantly less likely to have experienced major tsunami damage³.

The authors caution that vegetation does not prevent the worst devastation, but they see a protective role for plants in reducing damage from regular storms, such as the typhoons that batter the Philippines every year. These findings supported the locals' perceptions that mangroves are natural barriers, says Vaithilingam Selvam of the M. S. Swaminathan Research Foundation in Chennai, India. "It used to take us years to convince people that mangroves are worth saving," Selvam says. "Now it is much easier, because people see them as a bioshield."

Many southeast Asian governments have now pledged their support for mangrove restoration. Malaysia has promised \$25 million to replace 4,000 hectares of mangroves lost to the storm and to development. Indonesia has pledged \$22 million for mangroves, and has already planted 300,000 seedlings near the city of Banda Aceh. The Thai government has expressed its support for mangrove restoration and coastal rebuilding. In India, the government of the southern state of Kerala has pledged \$8 million to supplement an existing programme to restore mangroves destroyed by cyclones. International donor groups also offered money and seedlings, and several meetings were set up to coordinate the restoration projects.

Shifting ground

How likely are these projects to succeed? Many of the seedlings planted on the Aceh coast earlier this year have already died, because they were planted too soon and in the wrong places, say scientists working in the region. "Sometimes the site selection was not so good, and there was still debris from the tsunami washing up that destroyed the seedlings," says Faizal Parish, an ecologist and director of the Global Environment Centre in Selangor, Malaysia. And the paid contractors had no reason to continue caring for the young, fragile shoots.

Getting it right is slower and more difficult, say the groups with long-term experience of mangrove restoration in Asia. The Wetlands International Indonesia Programme (WIIP), based in Bogor, Java, has been running projects since 1998. The group coordinates replanting projects by negotiating agreements with villages. The villagers commit to plant a certain number of seedlings they collect from the wild. In return, the programme provides support staff and a loan that the villagers can use to start businesses, such as chicken and goat farms.

If 70% of the seedlings are still alive after five years, the village keeps the money; otherwise it must repay part of the loan. Nyoman Suryadiputra, a wetlands ecologist and technical director of WIIP, says this encourages villagers both to plant and care for the forests, making them more likely to survive. So far, villagers have replanted 350 hectares of land



Mucking in: involving local people is one of the keys to a successful restoration project.

along 3.5 kilometres of Indonesian coast through these arrangements.

But rebuilding after the tsunami has been more challenging, Suryadiputra says, and covers a wider area. Some of the group's current activities focus on the Aceh coast, including the project in Kajhu. But the group has also begun work on Simeulue Island, 300 km southwest of Banda Aceh, which was hampered by the tsunami and its aftershocks.

Before the disaster, Wetlands International estimated that Simeulue Island had at least 1,000 hectares of mangrove swamps. But the tsunami — plus another earthquake that followed this March — lifted parts of the island by 1–2 metres, cutting off mangroves and coral reefs from the tidal waters that sustain them. And the islanders have been so devastated by these disasters that — understandably — planting trees is not their first priority. "Many of the people are still traumatized," Suryadiputra says.

So the group first had to convince Simeulue



Under threat: southeast Asia has lost a quarter of its mangrove forest in the past 20 years.

residents that planting trees would restore some of their lost livelihoods. This summer, staff member Eko Budi Priyanto persuaded villagers in Langi, on Simeulue Island, to plant a nursery of 5,000 *Calophyllum* tree seedlings grown from seeds they collected on the beach. Building the nursery brought the village together, as entire families pitched in to help. The seedlings are now about a metre tall and have just been planted, Suryadiputra says. "It's been a really big challenge, and I don't want to say our approach will be successful," he says. "We want to do this properly from the beginning, and that's not easy."

Gathering storm

The time and effort involved in community-based programmes makes them harder to execute, but those with experience say approaches that involve local people are the only ones that make sense. In three states in India, where Selvam works, 33 villages have worked with forestry officials to restore 1,500 hectares of mangroves since 1993. So far, three-quarters of the seedlings have survived, Selvam says, double that of other techniques. The communities saw the benefits of their work when the trees buffered the impact of the tsunami, says Selvam, who is trying to enlist new villages in restoration projects.

The question is whether commitments to restoration will last longer than one year after the tsunami. Already, there are signs that governments and developers are returning to their old ways. In Indonesia, despite official support for buffer zones, shrimp farms destroyed in the tsunami are being rebuilt without setting aside space for trees. In some cases, says Parish, the Indonesian government has given newly damaged coastal land to developers, to create new shrimp farms. "These are strong pressures, as they were before," says Parish.

In many cases, local people are unhappy about these new developments, but they lack influence with government officials. This is the situation facing poor communities everywhere, not just in areas hit by the tsunami, says Jurgenne Primavera, a marine biologist at the Southeast Asian Fisheries Development Center in Iloilo, the Philippines. "The people who are at risk are poor, so they can't lobby the government to put protective greenbelts in the places where they live," says Primavera. The result in Indonesia, notes Parish, is that "the initial sensible plan to have some sort of greenbelt along the coast has not advanced as much as it could have."

The next storm in Kajhu is unlikely to be a tsunami, but without healthy mangroves it could still be devastating.

Erika Check is Nature's Washington biomedical correspondent.

1. Food and Agriculture Organization *State of the World's Forests* (FAO, Rome, 2003).
2. Naylor, R. L. *et al. Science* **282**, 883 (1998).
3. Danielsen, F. *et al. Science* **310**, 643 (2005).

BUSINESS

Olympus finds market rival hard to swallow

A video camera nestling inside a transparent capsule just a few centimetres long has put Japanese imaging giant Olympus on the back foot. Launched in 2001 by Israeli start-up firm Given Imaging, the capsule is a wire-free endoscope that is simply swallowed by the patient.

Famed for its cameras, Tokyo-based Olympus is also the world's largest manufacturer of conventional endoscopes. But despite claims that it has been working on the idea since the 1990s, the company unveiled its version of the capsule endoscope in November 2004 — and the product has so far only received approval for sale in Europe.

Given Imaging's PillCam was described by its inventor, Israeli missile engineer Gavriel Iddan, in *Nature* back in 2000 (G. Iddan *et al.* *Nature* **405**, 417; 2000). It offers a significant advantage over wire endoscopes as it can provide images from the whole length of the small intestine — a notoriously long, convoluted and difficult part of the gut to image. It also largely eliminates the patient discomfort associated with conventional endoscopies.

As it passes through the digestive system, PillCam can send back pictures for up to eight hours, and doctors have found it useful for diagnosing problems such as small tumours and bleeding. Some 260,000 patients across 60 countries have already been examined with the disposable probe, its maker says.

Those numbers have been good for business. In the nine months up to September this year, PillCam's sales reached US\$62.3 million worldwide, up 45.6% from the same period in 2004.

Left standing by the Israeli company, Olympus finally brought its capsule endoscope, called the Endo Capsule, to market in Europe this October. Clinical testing is still under way in the United States and Japan, and the company needs to receive full approval before it can begin selling its probe in these markets. Industry insiders add that Japanese firm RF System is developing similar products, but its looks as though it is still several years from the market.

Olympus's late entry into the market highlights the fact that Japanese companies tend not to capitalize on their innovations quickly, say analysts. "Many large Japanese companies don't feel the importance of speed and timing," says Yoko Ishikura, a specialist in corporate strategy

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

YOSHIKAZU TSUNO/AFP/GETTY IMAGES

Late launch: the Endo Capsule is coming to the market several years behind its competitor.

at Hitotsubashi University in Tokyo. "They try to play a game on their own turf, and are slow to look outside."

Iddan filed his first patent on PillCam in 1994 and helped to set up Yoqneam-based Given Imaging in 1998. The company now holds nearly 40 related patents. "PillCam has been so successful over the past five years that others cannot avoid trying to imitate it," claims Yoram Ashery, one of the company's vice-presidents.

Katsuyoshi Sasagawa, deputy general manager of Olympus Medical Systems, claims that his company was looking at capsule endoscopes throughout the 1990s. But it was 2003 before it marshalled its expertise in digital cameras and endoscopy to create a competing device. Olympus plans to pitch the price of the Endo Capsule close to the US\$450 charged by Given Imaging in the United States, and says it has been careful not to touch the smaller company's patent rights.

The Japanese company also claims that its product will have an edge thanks to its expertise in digital cameras, which should help it produce high-resolution images similar to those from conventional endoscopes. The PillCam currently uses slightly lower-quality imaging technology, but Given Imaging says it is working hard to improve this. As the competition between the two heats up, the struggle for supremacy could well go to a photo finish. ■

Ichiko Fuyuno

IN BRIEF

CANCER DEAL British biotechnology company Astex Therapeutics last week announced a deal with Swiss drug giant Novartis that is potentially worth US\$500 million. Under the agreement, Novartis will pay \$25 million to secure the global licensing rights to Astex's anticancer drug AT9311. The orally administered drug, which inhibits the cell cycle, is nearing clinical trials. Novartis has also obtained an option on the global licence for AT7519, an intravenously administered cell-cycle inhibitor, which is now in early clinical trials. Payments including fees, options and milestones could total US\$520 million, excluding royalties.

INDIAN EXPANSION Microsoft chairman Bill Gates said last week that the company will invest US\$1.7 billion and hire 3,000 workers in India over the next four years. The announcement came on the same day that the software company lost an anti-trust case in South Korea. As a result, Microsoft was fined US\$32 million and ordered to unbundle its Messenger and Media Player software from its Windows operating system. Despite the statement issued after the ruling in which the company reaffirmed its commitment to South Korea, Microsoft has previously said that losing the case could force it to withdraw Windows from the country. Microsoft's investment in India mirrors similar moves announced recently by Intel and Cisco Systems.

VACCINE MAKERS COURTED

The US Securities and Exchange Commission last week revised its interpretation of an accounting rule in an effort to encourage vaccine manufacturers to take part in strategic stockpile programmes. The change allows companies making vaccines for stockpiles such as pandemic influenza and biodefence to chalk up the vaccines as sales when they are placed in the stockpiles. Until now, firms had to wait until vaccines were distributed to claim the sales. Critics have argued that this may have discouraged companies from participating because stockpile supplies, by their nature, may never be used and have a limited shelf life.

Biodiversity: there's a role to be played by 'museum-keepers' too

SIR — Pierre L. Ibisch and colleagues, in Correspondence (“Biodiversity needs the help of global change managers, not museum-keepers” *Nature* **438**, 156; 2005), suggest that conservationists should not focus on single-species approaches in conservation science and practice. But the concept of ecosystem function may be difficult to explain to the general public. Many non-scientific conservationists I know would be more willing to donate money to save ‘cute’ or impressive animals, such as the Florida panther or the aptly named resplendent quetzal, than to support a theoretical science-based concept such as ecosystem services.

Valuing single species does not rule out using the services aspect of ecosystems. Take ecotourism, for instance: a ‘service’ as defined by the international Millennium Ecosystem Assessment. A Guatemalan farmer may not care about the resplendent quetzal if he is hungry and destroying more forest will give him more agricultural land. But if the farmer is offered a share in ecotourism projects organizing tours for bird-watchers who wish to see the resplendent quetzal, he may work to preserve the forests.

Many conservation organizations — both applied and science-based — need to focus on a single species in order to raise public money and support. But successful conservation probably needs to include ecosystem function and services to be successful in the long-term. Strategists should ask themselves: What can we do? What can we explain to the broad public? What can we explain to the locals? And what goals can we realistically achieve?

In the end, we need to be both global change managers and museum keepers.

Swen C. Renner

Conservation and Research Center,
Smithsonian Institution,
National Zoological Park, 1500 Remount Road,
Front Royal, Virginia 22630, USA

Climate research opponent is not a friend to science

SIR — In your Editorial “Taking a stand on animal-rights violence” (*Nature* **438**, 1; 2005) you suggest that US Senator James Inhofe (Republican, Oklahoma) should be applauded for supporting legislation that would make it easier to protect researchers against the threat of violence from animal-rights activists.

Some readers might get the impression

from your Editorial that Inhofe is a friend of science and scientific research. I strongly disagree with this.

This is the same Inhofe who has told the US Senate, and repeats on his website, that the threat of catastrophic global warming is the “greatest hoax ever perpetrated on the American people”. He has worked hard to discredit legitimate research in this area (see <http://inhofe.senate.gov/pressreleases/climateupdate.htm>) and has fought to exclude the findings of such research from public policy in the United States. He is co-sponsor of the Bush administration's energy bill (see <http://thomas.loc.gov/cgi-bin/query/z?c108:s485>) which you have previously — and justifiably in my opinion — characterized as inadequate (*Nature* **435**, 247; 2005). Inhofe has also taken strong stands against “all types of cloning” and has called embryonic stem-cell research “inconsistent and unreliable and ... unethical” (see inhofe.senate.gov/pressreleases/content-healthcare.htm).

Any calls to applaud Inhofe should be considered in the context of his broader — and in my view much more negative — record on scientific issues.

William P. Reich

Department of Psychology, Roosevelt University,
430 South Michigan Avenue, Chicago,
Illinois 60605, USA

Languages: Europe puts its money where its mouth is

SIR — Your News Feature “Tongue tied” (*Nature* **438**, 148–149; 2005), drawing attention to endangered languages, was very welcome, but in focusing on US research it overlooked the main work done in this area.

By far the largest investments in endangered-language research have been made by European institutions rather than US ones. Consider, for example, the ongoing investment by the Volkswagen Foundation (\$12.75 million and still rolling) and the Lisbet Rausing Charitable Fund (\$34 million invested through the School of Oriental and African Studies, London).

Even the research on the Siberian language Tofa that was reported in your News Feature was undertaken with Volkswagen funding: it is part of 5 terabytes of data now archived in www.mpi.nl/DOBES at the Max Planck Institute for Psycholinguistics.

In general, I believe that *Nature's* coverage of language and linguistics is relatively poor — a shame, as this is arguably the most advanced area of research in the humanities and it is developing fast.

Stephen C. Levinson

Max Planck Institute for Psycholinguistics,
PO Box 310, 6500 AH Nijmegen,
The Netherlands

Languages: land may speak louder than words

SIR — Your News Feature “Tongue tied” (*Nature* **438**, 148–149; 2005) shows that field linguists often attempt to save indigenous languages and the information embedded in them by gathering words into dictionaries and grammar texts. This process has an underlying assumption that words mean ‘things’, and that, once saved, they can be reassembled with the proper grammar to represent experiences.

This perspective is natural for people with a written language, but for many indigenous people, experience bears its own expression and names are ‘written’ in the terrain. Thus, saving a language may be more dependent on conserving the place in which the language arises.

This became apparent to me in a conversation I had with an indigenous woman from Alaska. Her language, Eyak, is also disappearing: only one old woman now has it as her primary language. While her speech is being recorded and children are encouraged to converse with her, this grandmother is not worried about the demise of Eyak. The elder's advice is to learn the common syntax and grammar of another Athabaskan language, such as Dine, and to return to the land where “the words will be available in the surroundings”.

Michael Stocker

Seven Circles Foundation, PO Box 559,
Lagunitas, California 94938, USA

Taxing question of when ethics go up in smoke

SIR — In News in Brief (“Research centre refuses tobacco-company funding” *Nature* **438**, 271; 2005), you report that the German Cancer Research Centre (DKFZ) has adopted “an ethical code that bans all kinds of funding from tobacco companies”. To me this sounds sanctimonious, because — as a member of the country's largest research institution, the Helmholtz Association — the DKFZ is funded almost entirely through taxes.

In 2003, the German federal government received €14.1 billion (US\$16.6 billion) in revenue from tobacco taxes, which amounts to 3.4% of the country's total tax income. As Dr Pötschke-Langer of the DKFZ hopes to “set an example for other German health institutes”, I wonder if the DKFZ aims to convince the Helmholtz Association to do without the €74.8 million that represents 3.4% of its annual €2.2-billion funding, or even to renounce 3.4% of its own budget?

Johannes Gerdes

Borstel Research Centre, Division of Tumour
Biology, Parkallee 22, D-23845 Borstel, Germany

BOOKS & ARTS

Changing your world view

Software that turns maps into landscapes reveals how cultural baggage can distort scientific images.

Landscapes Without Memory

by Joan Fontcuberta

Aperture Foundation: 2005. 96 pp. £22, \$40

Philip Ball

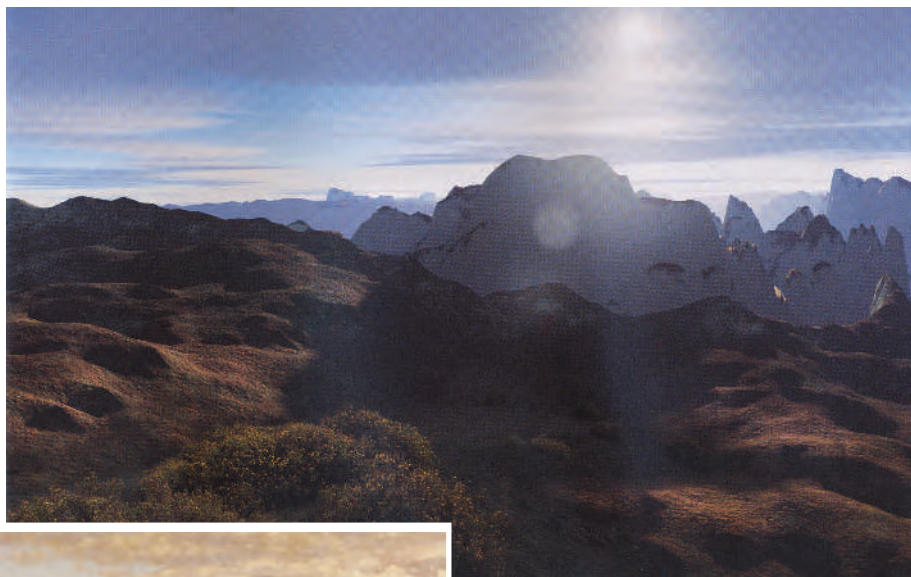
At first glance, Joan Fontcuberta's book of simulated landscapes looks like a straightforward attempt to extract a glossy coffee-table product from the marriage of science and art. Its padded cover, high-quality colour and landscape format add to the impression of indulgent luxury.

But it seems that the author's aim (if not, perhaps, the publisher's) is more postmodern. The 'art book' market seems to be the very arena that Fontcuberta, a Spanish artist and interpreter of the photographic image, wants at some level to challenge. The result is a perplexingly contradictory work — yet one that I think contains, whether intentionally or not, a potentially important message about visual representation in science.

This gallery of virtual landscapes has been generated by a fractal-based software package called Terragen, a tool used for turning maps into three-dimensional images that was originally developed for scientific and military applications.

Fontcuberta has subverted Terragen's purpose by feeding it 'false' data — not maps at all, but paintings of landscapes by famous artists and photos of his own body. By doing so, he demonstrates that Terragen can make a mountain out of anything (and will typically put a lake in the middle).

He points out that, although Terragen's advocates celebrate its 'realism' and its ability to capture the beauty of nature, running the program on the default settings gives a result that "tends to come very close to the kitsch of picture postcards". This is amply demonstrated in the book; indeed, in his introductory commentary, art historian Geoffrey Batchen speaks of the "relentless, banal, undemanding repetition of pictorial clichés" evident in these images, calling them "terrible in the way they give themselves up so easily to the demands of communal taste". In other words, he admits that, as art books go, this one is awful. And yet it is surely communal taste that will sell



A cultural landscape: the images produced by Terragen have the kitsch quality of the Romantic paintings of Frederic Edwin Church (left).

G. CLEMENTS/CORBIS

the book, whose readers may skip the text and simply bask in the 'beautiful' scenes that a computer has constructed from the raw pixels of paintings by Paul Cézanne, Georges Braque, Mark Rothko, André Derain and others.

Does Fontcuberta acknowledge this paradox himself? He hints, but doesn't quite come out with it. "Postmodernism, the society of the spectacle, the capitalism of fiction and this age of melancholy," he writes, have combined to consolidate "a mistrust of a reality composed of simulations, manifested in an avalanche of seductive, saccharine images — to which it is imperative that we respond critically." Is this an invitation to react with aesthetic horror to his pictures, to the way in which Terragen has turned great art (or biology) into a bland, kitsch vision of the Sublime? Or is he just hedging?

I think, however, that he may be identifying something more important than the mere fact that the computer geeks who created Terragen, probably weaned on bad sci-fi art, have

imbued their program with terrible taste. (Actually there is something rather enjoyable in letting your intuition interpolate between the 'map' and the Terragenized representation, noting how the colours and textures of Wassily Kandinsky and Thomas Gainsborough have been transmuted into these rugged slopes and brumal skies.)

The fact is that we have seen these landscapes before. Terragen has turned Braque's *The Fields* into something resembling Saturn's moon Titan, as recently unmasked by the Huygens probe, with its tawny skies reflected in petroleum lakes among ice-covered hills. Piet Mondrian's *The Grey Tree* has become red and barren Mars, and the heavy skies and golden peaks produced from Rothko are Venus as seen by the Magellan mission. It's a stark reminder that what NASA (or increasingly, as in Titan's case, an online army of sophisticated amateur graphics experts) is feeding us are visions of other worlds fed through a very particular stylistic filter.

In other words, these computer-generated images from space missions come laden with cultural baggage. The methods for generating these pictures from raw data have been imbued with what is probably a largely subconscious bias about how a landscape should look. And

it is not hard to discern the origin of that bias. Unquestionably, as Batchen points out, much of it stems from the German Romantics, particularly Caspar David Friedrich (whose *Wanderer Above the Sea of Fog* was, appropriately, Fontcuberta's first victim). But the form that this aesthetic takes in Terragen's creations is most clearly reminiscent of the American Romantics such as Thomas Cole (who was deeply influenced by Alexander von Humboldt's descriptions of the Andes) and his student Frederic Edwin Church. For generations of Americans, Church's heightened (in all

senses) landscapes were the way nature 'really' looked. That influence is clear in the planetary paintings that pre-date the current computer-generated versions, as well as in the vistas seen in any 'space opera' Hollywood movie.

But the influence of such artistic traditions on scientific imagery does not stop at real landscapes. Stock representations of the Sublime can also be discerned in company photographs of chemical plants soaring like tubular mountain ranges into coppery skies, and in the 'atomic landscapes' that can now be constructed from scanning probe microscopy. Even the

interpretation of the periodic table recently produced by Britain's Royal Society of Chemistry, with its crags thrusting out of a placid lake, has a Terragenic air.

More than 40 years ago, art historian Ernst Gombrich pointed out in his seminal book *Art and Illusion* (Pantheon, 1960) that no human depiction of the world (and this surely now includes photography) escapes our culturally acquired stereotyping. Maybe Fontcuberta's book will serve to remind scientists that the same applies to them too.

Philip Ball is a consultant editor for *Nature*.

Taking flight

Birds of South Asia: The Ripley Guide

by Pamela C. Rasmussen & John C. Anderton
Smithsonian Institution/Lynx Edicions: 2005.
Two volumes, 384 and 688 pp. £55

Richard Grimmett

S. Dillon Ripley is a titan of twentieth-century Indian ornithology, having written three major works: *Synopsis of the Birds of India and Pakistan* (Bombay Natural History Society, 1961), the ten-volume *Handbook of the Birds of India and Pakistan* (Oxford University Press, 1968–98), and *A Pictorial Guide to Birds of the Indian Subcontinent* (Oxford University Press, 1983), the last two in collaboration with Indian ornithologist Salim Ali. As secretary emeritus at the Smithsonian Institution, Ripley initiated a new bird guide for the Indian subcontinent, intended as his final major work on the region. To his credit, he engaged Pamela Rasmussen and John Anderton to work on the task. Ripley was taken ill shortly afterwards and died in 2001. But the project lived on and the long-awaited result has now been published.

The two-volume *Birds of South Asia: The Ripley Guide* covers the avifauna of the entire Indian subcontinent and includes 1,441 species. It expands on similar work by including Afghanistan and the Chagos Archipelago. *Volume 1: Field Guide* is portable and comprises 180 colour plates by John Anderton and other renowned bird illustrators, with brief adjacent text on field identification and distribution maps. The plates are generally good to excellent, with comprehensive coverage of plumage, although the illustrations of a few birds, such as the common nightingale, have suffered from an over-reliance on museum specimens rather than observations in the field. Several recently described species have been illustrated in a field guide for the first time here. *Volume 2: Attributes and Status* is a dense, comprehensive work that contains masses of new information on bird identification, variation, occurrence, habits, vocalizations and taxonomy.

The book's greatest value is that Rasmussen has taken nothing for granted, even informa-

tion published in Ripley's own works. Everything from bird distributions, measurements, vocalizations and identification features has been reviewed from scratch. The species list for the region has also been completely revised. Quite a few species are conservatively listed as 'hypothetical', with many previously published and significant records being regarded as 'insufficiently proven' (to the disappointment, no doubt, of many a living bird-watcher). Two well respected ornithologists from the first half of the twentieth century, E. C. Stuart Baker and Richard Meinertzhagen, are taken to task for their carelessness or fraudulent work (see *Nature* 437, 302–303; 2005). Their records — which underpinned Ripley's previous books on the subcontinent — are either treated with caution or dismissed.

Most significantly, and bravely, Rasmussen has given full species status to many forms for the first time in any modern guide. The common blackbird, for example, is treated as three species, with the Himalayan and south Indian forms elevated to full species. On this



Bird identification on a plate: magpies, jays and treepies, as illustrated in *Birds of South Asia*.

she is almost certainly correct, although readers will have to wait for further justification in the scientific literature before her judgements can be fully assessed.

Richard Grimmett is at BirdLife International Asia Division, Toyo-Shinjuku Building 2F, 1-12-15 Shinjuku, Tokyo 160-0022, Japan.

Tracing the history of art

Optics, Instruments and Painting, 1420–1720: Reflections on the Hockney–Falco Thesis [Early Science and Medicine Vol. 10 no. 2]

edited by Sven Dupré

Brill Academic: 2005. 214 pp. \$74

David G. Stork

Readers of *Nature* were among the first to learn of an intriguing theory proposed by artist David Hockney. He suggested that as early as 1420 some leading European artists used concave mirrors to project optical images on to their canvases, and traced them during the execution of their paintings (see *Nature* 400, 524 (1999); 412, 860 (2001); 417, 794 (2002)). The claim is sometimes called the Hockney–Falco thesis to acknowledge the technical efforts of physicist Charles Falco. Hockney thinks the procedure

was a key source of the naturalism, or 'optical look', arising in early Renaissance painting.

This special issue of *Early Science and Medicine* is a product of a four-day symposium in Ghent in November 2003. It is the first work to provide independent evidence on the material culture and documentary record, and analyses of optical knowledge, of the fifteenth century.

Did artists or scientists have access to suitable mirrors? Sara Schechner, curator of historical instruments at Harvard University, explores the state of optical fabrication at the time. She concludes that "Renaissance mirrors were far from offering the painter a short-cut to a detailed and naturalistic image of his subject." Early mirrors, she argues, "could not reflect or project clear, undistorted, 'photo-realistic' images, as Hockney and Falco suggest."

Renaissance art historian Yvonne Yiu of the University of Basel finds no surviving textual evidence for the procedure: “the silence of this considerable body of texts on the concave mirror projection method is deafening.” To Yiu it seems inconceivable that “well-informed contemporaries” would not have described a method that, according to Hockney and Falco, “revolutionized the art of their time”. Filippo Camerota, a historian of Renaissance scientific instruments at the Istituto e Museo di Storia della Scienza in Florence, points out that in the second half of the sixteenth century, when appropriate projections were first documented, Giambattista della Porta wrote: “If you are incapable of painting a portrait, this is a method you should know.” This is hardly support for the claim that tracing projections sparked a revolution in painting more than a century earlier.

The scientist Witelo’s *Perspectiva*, from about 1270, has previously been used as evidence that Renaissance scientists and possibly artists knew about appropriate projections. But Mark Smith, a historian of Medieval and Renaissance optics at the University of Missouri, provides an object lesson in the importance of expert knowledge when judging such a claim. Hockney and Falco take Witelo as saying that the image literally floats in space or is somehow projected to a location outside the mirror. “What Witelo really means,” says Smith, “is that the image will be located behind the reflecting surface”. The public might have concluded that Witelo was describing a projection of a real image and hence is closely related to Hockney’s method. A student of basic optics might have concluded that Witelo described a real image in space, not on a screen, and hence only distantly related to Hockney’s method. But after analysing the Latin text, context and contemporary thinking, Smith argues that Witelo was describing instead a virtual image, unrelated to Hockney’s method.

It seems that Hockney has recently retreated from his claim that artists actually traced projected images to a weaker view that artists merely saw and were influenced by such images. Philosopher of science Christoph Lüthy of the Radboud University of Nijmegen has read contemporary texts and concludes the former claim is “fairly implausible” and the latter “still awaits corroborating evidence”.

The volume’s editor, Sven Dupré, a historian of Renaissance optics, summarizes the conclusions of the contributors. It makes the Hockney–Falco theory “extremely unlikely as far as its application for the period before the first textual reference to image projection around 1550 is concerned”, he writes.

This well-written volume may close the door on the Hockney–Falco tracing thesis. But it should also prove a good resource, especially for optical scientists lacking a background in the history of art or science, from which to explore optics in the early Renaissance. ■

David G. Stork is chief scientist at Ricoh Innovations, Menlo Park, California 94028, USA.



Who am I? Alex (Elliot Levey) searches for a sense of self.

THEATRE

Self deception

On Ego

Soho Theatre in London, UK, until 7 January 2006.

Lucy Odling-Smee

What is this phenomenon that we call our sense of self? Is it the secular equivalent of the soul? Or is it just loops of wiring within the wet, grey stuff of our brains? Inspired by neuropsychologist Paul Broks’ book *Into the Silent Land* (Atlantic, 2003), theatre director Mick Gordon examines this dilemma in his new play *On Ego*.

The play opens with a lecture by a neurologist, Alex, who explains why the mass of neurons behind the face annihilates the myth of the soul. Equipped with slides of scalpels probing brain tissue, and a dripping human brain lifted from a bucket, he argues that the self is no more than an aggregate of an individual’s thoughts, feelings, perceptions and actions.

To make his point — that the ‘I’ is an illusion, and that we are no more than bundles of information — Alex enters a teleportation machine that vaporizes his physical body. The procedure involves scanning every atom of his body and transmitting the information by radio waves to a new destination. It should automatically eliminate the original Alex, but it goes wrong, leaving two versions, who immediately start acquiring different experiences and memories. When faced with the prospect of being eradicated, however, his original ‘self’ resists and an intuitive sense that he is an ‘I’ after all kicks in with a vengeance.

Meanwhile, the duplicate Alex is unaware

of his status and carries on as normal. But as the play proceeds he finds himself rejected by his wife, Alice, who is struck down with the psychological disorder Capgras’ syndrome — a condition that makes her believe that her husband is an impostor.

Physical duplication dances with illusion and delusion. Just as Alice’s perception of Alex changes, so does his perception of himself. The original Alex is confronted with two polar views of himself: his objective scientific view and his subjective view as he reacts to his experiences.

The many layers of the play, which stretches from the fantastical to a chilling portrayal of lives blighted by brain disease, is saved from chaos by being anchored to a carefully thought-out philosophical discussion about the nature of the self. While Alice struggles with her loss of self through lost love, her husband — and the audience — are left contemplating the space between what we rationally think we are and what we intuitively believe ourselves to be.

The stripping back of illusions to reveal the true self has long been the fodder of dramatists — think of King Lear peeling off his robes in the wind and the rain to reveal the “bare, forked animal” beneath. But in pushing against the boundaries of science, Gordon creates a tale that is both disturbing and curiously liberating. Perhaps we will never succeed in scientifically stripping back the intuitive sense of self, and perhaps, as Broks says, this is the “beautiful paradox” of being a human trying to understand itself. ■

Lucy Odling-Smee is a subeditor for *Nature*.

An asymmetric world

At the level of particles, things can happen in reverse, because particles obey time-symmetric laws of mechanics. But then why does matter, which is made up of these building blocks, behave irreversibly?

Oliver Penrose

My father once filmed a friend diving into a swimming pool. He liked to show the film in reverse, so that a big splash converged on the swimmer and expelled him from the water. The obvious impossibility of the reversed sequence of events illustrates the fact that our world is not symmetric under time reversal.

Most people are not worried by the asymmetry, but it does give physicists some trouble. Physicists want to explain the macroscopic behaviour of matter in terms of the microscopic mechanical laws obeyed by its component particles. These laws are symmetric under time reversal: for any motion obeying the laws of mechanics, the time-reversed motion also obeys these laws. Newton's laws have this symmetry, as do Maxwell's electromagnetic equations, Schrödinger's equation and Einstein's general theory of relativity.

So, naively, one would expect the matter made out of the particles obeying these laws to behave in a time-symmetric way. Yet it does not. Divers are not expelled from swimming pools by converging splashes even though there is a dynamically possible motion of the molecules in the pool that would produce this effect. Nor do natural processes such as heat flow ever go into reverse. The question is, then, how to explain the irreversibility of the macroscopic behaviour of matter, given the reversibility of the microscopic laws obeyed by its constituent particles.

This 'paradox of irreversibility' came to light when James Clerk Maxwell and Ludwig Boltzmann formulated their kinetic theory of gases. Boltzmann's own response (in 1905) was: "From the fact that the differential equations of mechanics are left unchanged by reversing the sign of time ... Herr Ostwald concludes that the mechanical view of the world cannot explain why natural processes run preferentially in a definite direction. But such a view appears to me to overlook that mechanical events are determined not only by differential equations but also by initial conditions."

Boltzmann's observation is certainly true, and many use it as the core of their explana-

tions of irreversibility. But such an explanation seems to beg the question. Why do we consider mechanical events to be determined by initial conditions, rather than by final conditions? Certainly, it is a fact of experience that you can start a simple mechanical system (a thrown ball, for example) in whatever state you choose, whereas the final state can be controlled, if at all, only by manipulating the initial state very accurately. But this special role of initial (as opposed to final) conditions in mechanics is another time-asymmetric feature of the world we live in — and would



Time's arrow: why can't you make a splash go backwards?

seem to require explanation along with the irreversibility of natural processes, rather than be used as part of that explanation.

The standard explanation of macroscopic irreversibility goes like this. Consider, as an example, a container separated into two equal-sized compartments by a partition. Initially one compartment contains gas and the other is empty. The partition is removed, and the gas expands irreversibly into the empty compartment. Soon it is distributed approximately evenly between the two compartments.

The explanation given is that the removal of the partition suddenly makes available to the system a much larger number of dynamical states (phase-space points). All but a tiny fraction of the dynamical states compatible with the initial macroscopic state (one compartment full, the other empty) lead to an approximately even distribution of gas between the two compartments, and therefore, so the argument goes, it is practically certain that the gas will end up approximately evenly distributed.

This argument is based on a probability assumption. The assumption is this: if practically all the possible initial dynamical states lead to a particular macroscopic behaviour, then it is practically certain that the actual system will exhibit that behaviour. The assumption is plausible enough, given our experience of such things as tossing coins. Only a tiny fraction of the ways of setting a spinning coin in motion will cause it to land on its edge, and our experience is that the coin is practically certain not to land on its edge.

But notice that use of future tense in "will exhibit". This signals another time-asymmetric feature of the world we live in, namely the time-asymmetric way that applied probability works. To my mind, this time asymmetry itself calls for explanation and should not lightly be incorporated into an explanation of the time asymmetry of something else.

The final stage in the standard discussion of the irreversibility paradox is to consider how the initial low-entropy state (a macroscopic state of tiny phase-space volume) was produced. It must have come from some even lower-entropy state, and so on back to the beginning of time, when the Universe was in a special macroscopic state of exceptionally low entropy. Some authors even seem to maintain that the low entropy of this primeval macroscopic state is, by itself, enough to account for the irreversibility of everything thereafter. But such a claim surely goes too far, as it takes for granted repeated applications of the asymmetric probability assumption just mentioned.

Moral: if you look at the world through rose-coloured spectacles, you cannot tell which parts of it really are rosy and which parts just look rosy. Likewise, if you use time-directed concepts such as initial conditions and applied probability as tools for looking at the irreversibility paradox, you risk arriving at an incomplete or even misleading solution to this complex puzzle. ■ Oliver Penrose is an emeritus professor in the Department of Mathematics at Heriot-Watt University, Edinburgh, and the author of *Foundations of Statistical Mechanics* (Dover, 2005).

R. KALVAR/MAGNUM PHOTOS

ESSAY

NEWS & VIEWS

ARCHAEOLOGY

Life on the Costa del Cromer

Wil Roebroeks

Flint fragments from eastern England constitute the earliest known evidence of human occupation of Britain. The climate was balmy, and the environment was home to a wide range of animals and plants.

About 700,000 years ago, Britain was connected to continental Europe, and the large rivers that drained central and eastern England meandered sluggishly into the North Sea basin. Sediments laid down by these lowland rivers are found today along the coastline of northern Suffolk and Norfolk. As the sediments were deposited, remains of animals and plants became trapped in them: large and small mammals, reptiles, molluscs, and even trees, fruits and seeds, after which the Cromer Forest-bed Formation was named. Parfitt *et al.* (page 1008 of this issue¹) show that, along with hippos, rhinos and elephants, early humans were evidently roaming the banks of these rivers. They did so during a warm interglacial period, and much earlier than hitherto thought for this part of Europe.

Charles Lyell, the Victorian geologist, would have been pleased with Parfitt and colleagues' report — in the 1860s he had predicted² that evidence of human occupation would show up in the Cromer Forest-bed (the well-exposed deposits were then considered to be more or less of the same age as sediments already yielding early artefacts on the other side of the English Channel, in France). But despite the efforts of generations of fossil hunters, no unambiguous evidence was found until the discoveries reported by Parfitt *et al.*, which come from ancient river sediments exposed near the village of Pakefield, Suffolk. Much of Pakefield village as Lyell would have known it has been lost to the sea through coastal erosion of the cliffs, a serious problem in this part of Britain. But what can be a disaster for the home-owner represents an opportunity for the collector of fossils and artefacts.

Archaeologically speaking, Parfitt and his colleagues¹ have struck Stone Age gold. From part of the Forest-bed, they have recovered clear evidence for the presence of humans at an estimated 700,000 years ago. Working with intense effort at low tides, they have excavated 32 pieces of worked flint from the exposures along the shoreline near Pakefield (Fig. 1). They have also gathered together a remarkably rich set of data on the interglacial environment of these early pioneers.

Natural processes acting on flints sometimes



Figure 1 | The cliff at Pakefield on the Suffolk coast. The Cromer Forest-bed Formation investigated by Parfitt *et al.*¹ — the lowermost dark band — is overlain by thick glacial deposits.

produce artefact-like forms. Especially in the early days of their discipline, prehistorians often had great difficulty in telling the difference between such pseudo-artefacts and pieces of flint modified by human agency. Indeed, in the early 1900s the Forest-bed exposures yielded controversial primitive flint 'tools', which were promoted by some as evidence of early occupation³ but were eventually debunked as natural products by others⁴. Archaeologists have learned from such debates, and the Pakefield evidence for human activity is rock solid. The small assemblage consists mostly of waste flakes produced during flint knapping.

The assemblage is too small to be representative of the whole range of tools that was probably produced, and it is useless to speculate on the technological capacities of its makers. But I would not be surprised if some of the sharp edges revealed microscopic evidence of their former use as butchering tools, because animal products must have been part of the human diet in that environment⁵.

Small and simple though the Pakefield set of artefacts is, for archaeologists it has grand implications. Hitherto, the earliest unambiguous

traces of a human presence in Europe north of the Alps were dated to about half-a-million years ago⁶, and included the spectacular finds from Boxgrove, on the southern coast of England. There, thousands of flint artefacts have been excavated, together with the bones of butchered large mammals and even some human remains^{7,8}. The earliest traces of human presence in southern Europe — for example the rich materials from Atapuerca in Spain⁹ — are at least 800,000 years old (Fig. 2, overleaf). The southern European evidence suggested that there had been a long time-lag between the first occupation of this zone and that of the more northern parts of Europe, with humans being confined to the Mediterranean perimeter and Spain for a few hundreds of thousands of years before moving into the north.

The work at Pakefield has now shown this view to be flawed. No matter that one can quibble about the intricacies of the dating arguments for Pakefield (and I am sure people will), the site is probably 200,000 years older than any known previously from Europe north of the Pyrenees and Alps. In that sense, the discoveries come as a big surprise. In another sense,

S. A. PARFITT

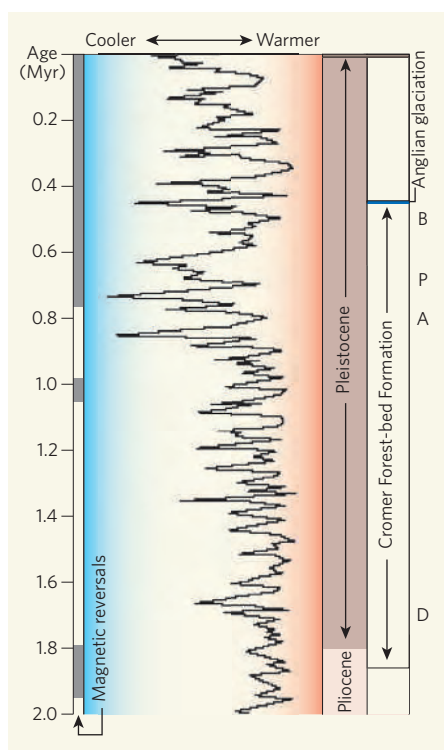


Figure 2 | Dating the human occupation of Europe. This chronology (in millions of years) shows the incidence of reversals in the geomagnetic field, and the cooler and warmer intervals as judged by oxygen-isotope studies, which provide benchmarks for dating. The Cromer Forest-bed Formation is capped by glacial deposits of the Anglian glaciation. The Boxgrove site (B) dates to the temperate period before the Anglian; Pakefield¹ (P) is at least two interglacials older. The Spanish site Atapuerca (A) is older than Pakefield, dating to just before the period of 'normal' geomagnetic polarity that started about 800,000 years ago. All of these sites are much younger than that at Dmanisi in Georgia (D), which is situated in the Caucasus at the 'gates of Europe'. This site yielded a rich fossil collection of small-brained humans dating to about 1.7 million years ago.

however, Pakefield fits well into earlier views of the colonization of Europe^{5,10,11}. As Parfitt *et al.*¹ point out, the environmental context of the flint assemblage provides a good explanation for the presence of humans in northern Europe: judging from the rich palaeoecological and climatic data from Pakefield, the range of these pioneers expanded temporarily in parallel with an expansion of their familiar warm, Mediterranean-like habitat. The Pakefield artefacts probably do not testify to a colonization of the colder temperate environments of northern Europe, but more to a short-lived human expansion of range, in rhythm with climatic oscillations. Although they occur in England, the finds are basically still 'Mediterranean' in that they were produced along the balmy shores of what can be seen as an early Middle Pleistocene Costa del Cromer. As in Asia¹², more significant occupation of the northern (colder) parts of Europe did not

occur until later, maybe from the times of the Boxgrove *Homo heidelbergensis* population onwards⁵. But the sea continues to expose long-buried sediments, and in due course more surprises may turn up — especially now that Parfitt *et al.* have finally demonstrated the archaeological potential of the Cromer Forest-bed.

Seen yet another way, however, the Pakefield evidence is just plain tantalizing. The Cromer Forest-bed is among the best-studied Pleistocene exposures anywhere in the world: it is a place where many generations of scientists have kept a sharp eye out for traces of early humans^{3,4}. If chronological surprises can turn up in such a location, what are the implications for other parts of Europe and of the world that have seen less and sometimes no archaeological attention? The finds from Pakefield will surely influence our understanding of the human occupation of Europe. But especially on a global scale, they are a reminder that we must

be terribly careful with translating absence of evidence into evidence of absence.

Wil Roebroeks is in the Department of Archaeology, Leiden University, PO Box 9515, 2300RA Leiden, The Netherlands.
e-mail: w.roebroeks@arch.leidenuniv.nl

1. Parfitt, S. A. *et al.* *Nature* **438**, 1008–1012 (2005).
2. Lyell, C. *The Geological Evidence of the Antiquity of Man, with Remarks on Theories of the Origin of Species* (John Murray, London, 1863).
3. Reid Moir, J. *The Antiquity of Man in East Anglia* (Cambridge Univ. Press, 1927).
4. Coles, J. M. in *Studies in Ancient Europe* (eds Coles, J. M. & Simpson, D. D. A.) 17–43 (Leicester Univ. Press, 1968).
5. Roebroeks, W. *J. Hum. Evol.* **41**, 437–461 (2001).
6. Dennell, R. & Roebroeks, W. *Antiquity* **70**, 535–542 (1996).
7. Roberts, M. B., Stringer, C. B. & Parfitt, S. A. *Nature* **369**, 311–313 (1994).
8. Roberts, M. B. & Parfitt, S. A. (eds) *Boxgrove: A Middle Pleistocene Hominid Site at Eartham Quarry, Boxgrove, West Sussex* (English Heritage, London, 1999).
9. Carbonell, E. *et al.* *Science* **269**, 826–829 (1995).
10. Dennell, R. *J. Hum. Evol.* **45**, 421–440 (2003).
11. Gamble, C. *The Palaeolithic Societies of Europe* (Cambridge Univ. Press, 1999).
12. Dennell, R. *Asian Perspect.* **43**, 205–226 (2004).

FLUID DYNAMICS

Drat such custard!

Troy Shinbrot

The mixing of festive sweetmeats and the stirring of cream into coffee are toothsome examples of the irreversibility of physical processes. In certain systems, however, the concept gets its just desserts.

Running a film sequence in reverse is a joke used to great effect in early moving pictures: the Keystone Cops round a corner backwards on two wheels and fall back into their wagon; collapsed buildings reassemble; and water climbs up a mountainside. The humour, and surprise, of these actions arises because we readily detect that time is travelling in the wrong direction. Similarly, when we stir cream into coffee, we know the two will mix; equally, were we to reverse the direction of stirring, we would be startled to see the coffee and cream return to their original, separated state.

On page 997 of this issue, Pine and colleagues¹ establish exactly when and where this counterintuitive phenomenon can occur. They examine suspensions of simple spheres in a liquid, and in their supplementary information supply graphic video evidence of a transition between a reversible flow (which returns to its original state when stirring is reversed) and an irreversible flow (which continues to mix regardless of the direction it is stirred in). Suspended particles can thus, like the Keystone Cops, exactly retrace their steps as the direction of stirring is reversed, or alternatively can diffuse irreversibly over time — as experience teaches us to expect. The transition between a reversible and an irreversible flow depends both on the concentration of solid matter and on how far the liquid has been

stirred: for low concentrations of solids stirred short distances, mixing can be reversed, whereas for higher concentrations stirred further, it cannot.

For their demonstration, Pine *et al.* used the flow of a fluid held between two concentric rotating cylinders. This system has had a long history in investigations of reversibility, ever since a celebrated demonstration in 1966 by the renowned fluid dynamicist and educator G. I. Taylor. Taylor injected a spot of dye into glycerine trapped between two concentric cylinders, and mixed one into the other by rotating the inner cylinder clockwise. Then, startlingly, by rotating the same cylinder anticlockwise, he was able to return the dye to its original state.

The flow between rotating cylinders has since been scrutinized numerous times^{2–5} in attempts to understand when suspended or dry grains behave like a solid, bouncing off one another in an irreversible manner like ping-pong balls, and when they behave like Taylor's viscous fluid, returning to their original positions when the cylinder rotations are reversed.

Pine *et al.* show¹ that both reversible and irreversible flows can be obtained predictably in such a system. Moreover, they demonstrate that irreversibility is directly associated with the occurrence of multiple collisions between

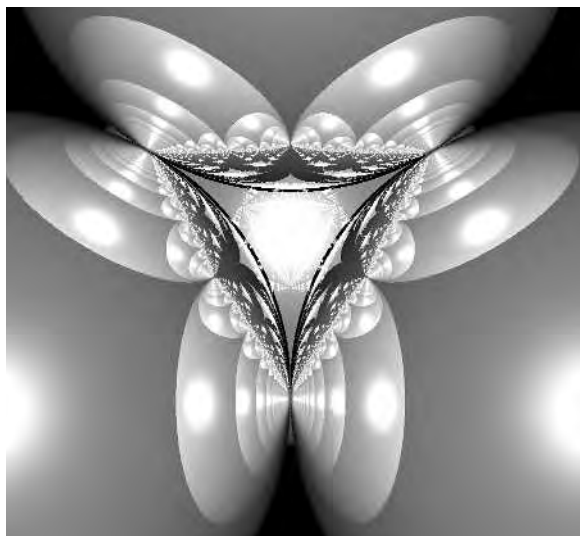


Figure 1 | Mirror, mirror on the ball. Intricate interleaved domains of scattering can be seen graphically in this computer-generated picture of three reflective spheres. In this system — easily reproduced by placing three Christmas-tree ornaments in contact — light scatters from one convex surface to another to create a ‘hall of mirrors’ effect: images of the light reflected from any one of the three balls can be seen in infinitely fine detail on the other two. This phenomenon of chaotic scattering is also seen in the system of suspended particles described by Pine *et al.*¹, where it leads to a transition between reversible and irreversible flows.

suspended particles. Explicitly, when the relative distance travelled by neighbouring suspended particles is small, the particles move back and forth along the same, unchanging, trajectories. By contrast, when the particles’ relative motion goes beyond a certain threshold that depends on their concentration, multiple particles interact and so diffuse irreversibly. Reducing the particle concentration makes multiple particle encounters much rarer and weaker, and can transform an irreversible flow into a reversible flow at a predictable point.

These results shed light on many practical problems, including the formulation of pharmaceutical suspensions, the catalysis of petrochemicals in fluidized beds, and the culture of shear-sensitive haematopoietic (blood-making) cells. In all of these applications, particles are suspended in fluids; yet, with rare and limited exceptions, mixing and flow in these suspensions cannot be predictably scaled up from the laboratory bench (typically in milligram quantities) to the production plant (in quantities up to millions of tonnes). This inability stems from our lack of a fundamental understanding of the interplay between microscale and macroscale physics in complex suspensions. Pine and colleagues’ work¹ indicates that this interplay may depend, above all, on the dynamics of collisions between the suspended particles.

As so often happens, this practical result was presaged by abstract mathematical modelling from decades earlier⁶. In this scheme, the intimate relationship between particle collisions on the microscale and mixing behaviours on the macroscale is caused by chaotic scattering between suspended particles: when suspended particles collide, they emerge from the collision in a chaotic direction that cannot be simply reversed to recover the original trajectories. Chaotic scattering has in turn been linked to baroque, interleaved ‘domains of attraction’ that emerge in widely differing systems where multiple bodies interact⁷, such as nonlinear pendulums, multi-mode lasers,

fusion-reactor plasmas, the orbits of celestial bodies and many others (Fig. 1). The intriguing interrelation between these diverse systems is largely responsible for the historical

richness of the simple problem of mixing between concentric cylinders.

Tradition has it that the stirring of mince pies and plum puddings during this holiday season must be done clockwise for good luck. Armed with Pine and colleagues’ results, we may be forgiven for indulging this primitive superstition: counter to the expectation imposed by strict logical reasoning, the preparation of pies and puddings can indeed depend on the direction of stirring. ■

Troy Shinbrot is in the Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey 08854, USA.

e-mail: shinbrot@soemail.rutgers.edu

1. Pine, D. J., Gollub, P., Brady, J. F. & Leshansky, A. M. *Nature* **438**, 997–1000 (2005).
2. Gadala-Maria, F. & Acrivos, A. *J. Rheol.* **24**, 799–814 (1980).
3. Mueth, D. M. *et al.* *Nature* **406**, 385–389 (2000).
4. Hartley, R. R. & Behringer, R. P. *Nature* **421**, 928–931 (2003).
5. Conway, S. L., Shinbrot, T. & Glasser, B. J. *Nature* **431**, 433–437 (2004).
6. Sinai, Ya. G. *Russ. Math. Surv.* **25**, 137–189 (1970).
7. Sweet, D., Ott, E. & Yorke, J. A. *Nature* **399**, 315–316 (1999).

NEUROSCIENCE

A painful factor

Carole Torsney and Amy B. MacDermott

Peripheral nerve injury activates cells in the spinal cord called microglia. But how do such cells cause the ensuing chronic pain? It seems that they release a small protein that disrupts normal inhibition of pain signalling.

Neuropathic pain is a debilitating chronic pain condition with limited treatment options¹. It is caused by damage to the nerves that transmit sensory information (touch, pain, temperature and so on). Patients experience crippling pain in response to stimuli that are not normally painful — a condition termed allodynia — and also suffer spontaneous and exaggerated pain. Research on neuropathic pain has tended to focus on injury-induced changes in sensory nerves and secondary changes in the spinal-cord neurons that receive and process sensory information before it is relayed to the brain to be ‘perceived’. But recent work shows that nerve injury also activates immune-like scavenger cells — microglia — in the spinal cord and implicates these cells in the resulting neuropathic pain².

Now Coull *et al.* (page 1017 of this issue)³ identify a critical link between activated microglia and the altered sensory-neuronal processing that underlies neuropathic pain. They show that activated microglia disrupt the inhibitory control of key spinal-cord pain neurons, and — crucially — they find that a small neuronal modulating protein called BDNF (for brain-derived neurotrophic factor) mediates this signalling between microglia and neurons.

Coull *et al.* have previously shown⁴ that

neuropathic pain is linked to injury-induced alterations in neurons in the spinal cord’s lamina I region, where some pain pathways to the brain originate⁵. Their work examined the action of the neurotransmitters glycine and GABA, which normally dampen neuronal excitability. Both transmitters are released from ‘inhibitory’ neurons, and then bind to receptors on the surface membrane of target neurons. The GABA_A and glycine receptors are also ion channels, which open when activated, allowing anions such as chloride (Cl[−]) to flow down their electrochemical gradient into the neuron. This is because the potential at which the current carried by anions reverses (E_{anion}) is negative with respect to the neuron’s resting potential (V_{rest}) (Fig. 1a, overleaf). The influx of anions makes the neuron’s membrane more negative, or hyperpolarized, relative to the resting membrane potential. Hyperpolarization then acts to limit neuronal activity, in this case inhibiting the lamina I pain pathway.

Using a rat model, Coull *et al.*⁴ showed previously that nerve injury somehow reduces the levels of the potassium-chloride co-transporter KCC2 in the membranes of spinal-cord neurons, so that the intracellular chloride concentration increases. This shifts E_{anion} so that it is positive with respect to the resting membrane



50 YEARS AGO

A striking picture of the dangers which confront men of science has been prepared recently by Gerard Piel... At the outbreak of the Second World War, he writes, science was a kind of ideal world republic. The scientific community was an international community. It was the only truly international community at an epoch that was to see nationalism and the narrower concerns of national power rise again to ascendancy in world politics. Statesmen discovered early in the War that science is an essential element of national strength... Accordingly, each major power has sought to monopolize the talents of its scientists and to put them to work in the name of national security. This suppression of international motives in favour of national ends has now had serious consequences upon the life of science... Scientists are agreed, in the first place, that there has been a dangerous diversion of resources and talent from the really significant long-range concerns of science to the narrower short-range objectives of practical results.

From *Nature* 17 December 1955.

100 YEARS AGO

The report of the chief of the United States Weather Bureau for the fiscal year 1903–4 contains... an interesting account of the very useful operations of that organisation. Weather forecasts for thirty-six and forty-eight hours in advance are issued for each State, besides special warnings of gales, cold waves, floods, &c. To mention one case only of the utility of storm warnings — a hurricane which advanced from the West Indies destroyed property to the value of 100,000 dollars during its progress over Florida, but, owing to timely notice, comparatively little damage was done to vessels, as they remained in port in consequence of the warnings. Prof. W. L. Moore reiterates the hope that the time will come when it will be possible to forecast the weather for coming seasons, but that time has not yet arrived.

From *Nature* 14 December 1905.

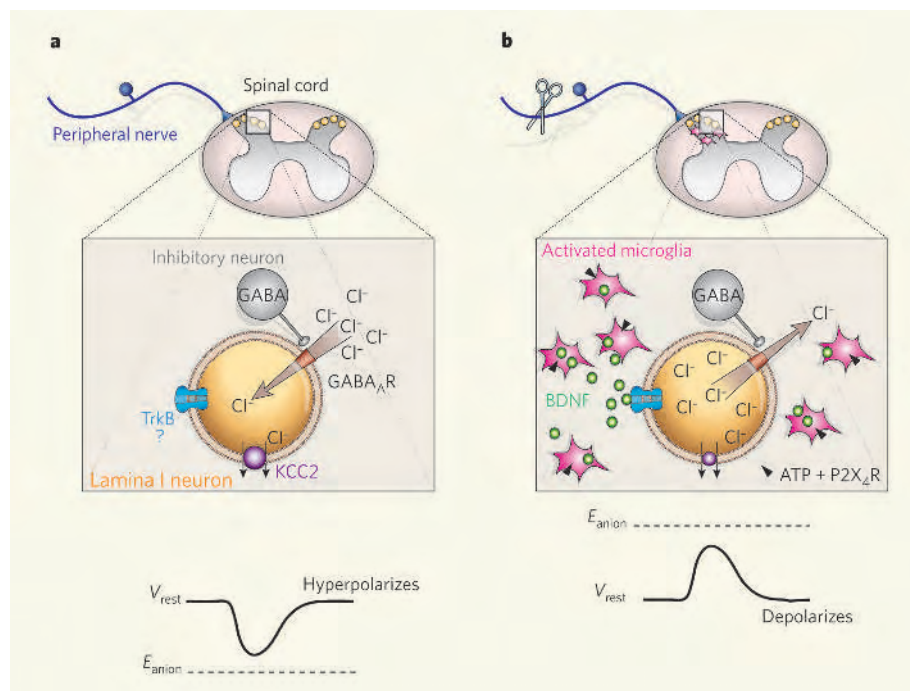


Figure 1 | A pathway of pain. Coull and colleagues³ reveal that microglial–neuronal signalling mediated by BDNF disrupts inhibition of rat lamina I spinal-cord neurons and maintains neuropathic pain. **a**, Activation of GABA_A receptors (GABA_AR) normally leads to an influx of anions (principally chloride, Cl⁻), causing hyperpolarization (inhibition), because the potential at which the anion flux switches from inward to outward (E_{anion}) is negative with respect to the resting membrane potential of the neuron (V_{rest}). **b**, Following peripheral nerve injury, activated microglia (with ATP-stimulated P2X₄ receptors) release BDNF, which acts on the TrkB receptor to modify E_{anion} , probably by reducing levels of the potassium–chloride co-transporter KCC2. As E_{anion} is now positive with respect to V_{rest} , GABA_A-receptor activation leads to an efflux of anions, depolarizing the lamina I neurons. Blockade of this microglial–neuronal signalling pathway alleviates chronic neuropathic pain in the rat model.

potential. GABA_A-receptor activation then results in anions flooding out of the neuron, making the neuron's membrane potential more positive, or depolarizing, relative to the resting membrane potential (Fig. 1b)⁴. As a result, the normally inhibitory transmitters GABA and glycine are no longer able to suppress signalling in the lamina I pain pathway.

In the latest study³, Coull and colleagues investigated whether the activation of microglia following nerve injury is responsible for this switch in the effects of glycine and GABA. Increased synthesis and activation of the ATP receptor P2X₄ found on microglia are required for the development of allodynia following nerve injury². The authors therefore stimulated microglia with ATP and applied them to the spinal cord of rats. The animals developed allodynia, as assessed by paw withdrawal from a light mechanical stimulus. Electrophysiological recordings in spinal-cord slices taken from these allodynic animals showed that ATP-stimulated microglia positively shifted the E_{anion} in lamina I neurons and rendered GABA effects depolarizing, rather than hyperpolarizing, in these neurons.

How do activated microglia communicate with lamina I neurons? To answer this question, Coull *et al.* examined the effects of BDNF, as this protein is secreted by microglia⁶, is involved in chronic pain⁷ and can cause shifts in anion gradients in the brain⁸. The authors

found that spinally administered BDNF produced allodynia and induced the predicted change in the anion gradient, enabling GABA to depolarize the lamina I neurons rather than inhibit them. In a rat model of nerve injury, blocking the spinal action of BDNF on its receptor, TrkB, reversed an established allodynia, confirming that BDNF is actually released in the spinal cord and is required for the development of neuropathic pain. Furthermore, when BDNF–TrkB signalling was prevented, spinal slices from nerve-damaged animals did not show the typical depolarizing shift in E_{anion} .

Coull *et al.* then confirmed that microglia are indeed the source of the BDNF, which until now was thought to be released from neurons during pain processing. They showed that ATP-stimulated microglia were unable to produce allodynia or shift E_{anion} when spinal BDNF–TrkB signalling was blocked. The authors also created microglia that could not synthesize BDNF, and showed that, although these microglia have otherwise normal ATP responses, they did not release BDNF when stimulated with ATP. Significantly, these BDNF-deficient microglia did not cause allodynia or shift E_{anion} when administered spinally.

This work firmly establishes BDNF as a crucial mediator of microglial–neuronal signalling during neuropathic pain. Furthermore, it highlights several remaining questions. Where does the ATP that stimulates

microglia come from? What signalling pathway leads to the release of BDNF from microglia? How does BDNF–TrkB signalling alter E_{anion} in lamina I neurons — could it perhaps lead to reduced synthesis of the KCC2 chloride transporter in the same neurons, or are TrkB receptors activated on other spinal cells that then influence lamina I neurons? Finally, what is the underlying circuitry that mediates the sensation of neuropathic pain when the actions of GABA and glycine are disrupted in the lamina I pain pathway?

Coull and colleagues' results provide an optimistic outlook for the treatment of neuropathic pain, because disrupting BDNF signalling was able to reverse established allodynia in the rat model. This suggests that continuous microglial–neuronal signalling is

required to maintain allodynia, and that it may be possible to treat the condition even after the neuropathic pain state is established.

Carole Torsney and Amy B. MacDermott are in the Department of Physiology and Cellular Biophysics, Center for Neurobiology and Behavior, Columbia University, New York, New York 10032, USA.
e-mail: abm1@columbia.edu

- Scholz, J. & Woolf, C. J. *Nature Neurosci.* **5**, 1062–1067 (2002).
- Tsuda, M. et al. *Nature* **424**, 778–783 (2003).
- Coull, J. A. et al. *Nature* **438**, 1017–1021 (2005).
- Coull, J. A. et al. *Nature* **424**, 938–942 (2003).
- Mantyh, P. W. & Hunt, S. P. *Trends Neurosci.* **27**, 582–584 (2004).
- Nakajima, K. et al. *J. Neurochem.* **80**, 697–705 (2002).
- Thompson, S. W. et al. *Proc. Natl Acad. Sci. USA* **96**, 7714–7718 (1999).
- Rivera, C. et al. *J. Cell Biol.* **159**, 747–752 (2002).

climatic and tectonic events sealed in these deep-sea sediments. On the one hand, measurements of the volumes of these deposits allow variations in the erosion rates of the Asian mountain belts over time to be estimated. On the other, distinctive chemical fingerprints in the sediment permit an evaluation of where the sediment came from in the first place. A combination of the two methods yields a picture of a highly dynamic hinterland in terms of tectonics, evolving topography and river discharge.

Clift and Blusztajn¹ reconstruct the discharge of particulate sediment of the ancient Indus River over the past 30 million years using data obtained from seismic surveys of the sediment on the floor of the Arabian Sea. During this time, the collision of India with Eurasia has formed the mountains of the Himalaya, a bountiful source of sediment for the rivers draining into the ocean. The chemical fingerprint of the sediments of the Indus fan is the rare-earth element neodymium⁶. This is expressed as ϵ_{Nd} , which is related to the ratio $^{143}\text{Nd}/^{144}\text{Nd}$ compared with a standard for the Earth, and is sensitive to the type of source rocks.

The currently low values of ϵ_{Nd} in modern Indus River sands appear to be part of a trend that started after five million years ago. Strongly negative ϵ_{Nd} values, typical of the sediments of the Bengal fan, are thought to be supplied by sources in the Himalayan mountain ranges marking the frontal zone of crumpling between India and Asia. More positive ϵ_{Nd} values are associated with source areas located behind the Himalaya, particularly in the Karakoram Range of northeast Pakistan, which is drained by the headwaters of the Indus. So Clift and Blusztajn make the intriguing connection that

EARTH SCIENCE

Volte-face in the Punjab

Philip A. Allen

Rivers are the great conveyor belts that carry sediment from mountains to the sea. In the Punjab — the Land of Five Rivers — a wholesale shift occurred in the past that re-routed sediment to different oceans.

Rivers don't come much bigger than the Ganges and the Indus, both of which drain the mighty Himalaya. However, as Clift and Blusztajn¹ show in this issue (page 1001), size does not mean permanence. Around five million years ago, the rivers of the Punjab evidently shifted from flowing into the Ganges system and the Bay of Bengal to flowing via the Indus system into the Arabian Sea. This major diversion of continental drainage has been deciphered from the isotopic signature of minerals collected from the Indus fan, a vast undersea cone of sediment stretching for more than 1,000 kilometres from the mouth of the Indus River.

It is well known that rivers shift their courses — switching of the position of the main channel within a river valley is historically well documented, and typically takes place at intervals of decades to thousands of years. Many of the lowland tracts of the world's major rivers, which flow through some of the most densely populated parts of the Earth's surface, and which were the sites of long-since-disappeared civilizations, contain a remarkable record of such switches. The Po of northern Italy and the Huang He (Yellow River) of eastern China are excellent examples. But what is less well understood is the wholesale shifting of river courses at the longer timescales described by Clift and Blusztajn.

The total amount of sediment discharged into the world's ocean is about 20 billion tonnes per year^{2–4}, and a high proportion of this global annual budget comes from the river systems of southeast Asia, from the Indus to

the Huang He⁵. The sediment dumped onto the sea bed of the Indian and western Pacific oceans over the past tens of millions of years has created a vast apron, a few kilometres thick near the mouths of the major rivers, which gradually thins seawards over distances of several thousand kilometres. Scientists are increasingly piecing together the record of past



Figure 1 | Drainage of the Himalaya. The map shows the configuration of rivers as it is today, with the rivers of the Punjab flowing into the Indus and delivering sediment into the Indus fan in the Arabian Sea. According to Clift and Blusztajn's¹ analysis of neodymium isotopes in sediments, however, before about five million years ago the Punjabi rivers instead flowed southeast (see Fig. 1b of the paper¹ on page 1002), delivering sediment into the Ganges and the Bengal fan. The growth of the Salt Range is one possible contributory factor to this diversion. Others are flexing of the Indian tectonic plate in response to mountain building and erosion, and the effects of climate change on river discharge.

MICROBIOLOGY

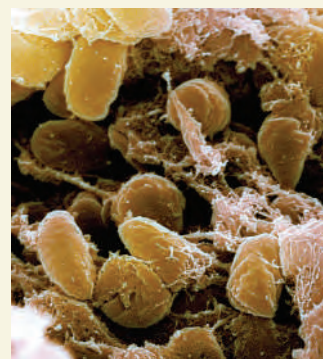
Perspectives on plague

The bacterium *Yersinia pestis* (pictured) is notorious as the cause of bubonic plague. When it is breathed in, however, it also causes the rarer but deadlier pneumonic plague. The pathology of this disease in humans and animals is fairly well understood, but much less is known about the earliest stages. Wyndham W. Lathem and colleagues (*Proc. Natl Acad. Sci. USA* **102**, 17786–17791; 2005) have developed a mouse model of pneumonic plague that gives perspectives on these stages as experienced by the host and by the bacterium.

The team infected mice with *Y. pestis* through the nose, and the animals developed a disease that closely resembled pneumonic plague in humans. Bacterial numbers in the animals' lungs increased massively in the first 24 hours after infection. Yet when the authors studied the levels of inflammatory molecules normally produced during an immune response, there was little change during this period. So the bacteria must have a potent anti-inflammatory activity that allows them to become established before the host immune system

detects them. After 48 hours, the levels of inflammatory molecules escalated, showing that the mouse immune response does eventually kick in; but it would seem to be too little, too late.

And what happens in *Y. pestis*? A microarray analysis showed that there is a change in the expression of about 10% of the bacterium's genes after it infects its host. Notably, many of these genes are associated with virulence, and in particular with the so-called type III secretion system. This system was already known as a potential means for the bacterium to subvert its host's immune system by altering the types and amounts of inflammatory molecules. That the expression of genes for the system



NIAD/CDC/SPL

is increased in the mice confirms this animal model as biologically valid. Moreover, a comparison with *in vitro* studies showed that the regulation of this system is more complex *in vivo* — suggesting that the model will provide greater insight into this devastating infection. **Helen Dell**

before five million years ago the rivers of the Punjab flowed eastwards as part of the Ganges system to feed the Bengal fan (Fig. 1).

This explanation of the former drainage of the Land of Five Rivers makes sense in the light of the sediments deposited at the foot of the Himalaya, known as the Siwalik Group sediments. These were deposited by ancient river systems that cut into the rising Himalayan mountains, and indeed it was previously suggested^{7,8} that there was a continental-scale flip-flop of drainage between the Bengal and Indus sinks. The real value of these new results¹ is therefore not the idea that drainage diversion can occur on a continental scale, but that the isotopic data from the deep-sea Indus fan provide such striking support for that view.

We are now developing a more dynamic impression of the way in which sediment is routed from mountains to the sea. Such routing is strongly influenced at a relatively local scale by the emergence of new mountain ranges in response to continuing continental convergence. For example, the growth of the Salt Range of northern Pakistan may have triggered the diversion of the main tributaries of the ancient Indus to the Arabian Sea. But at the larger scale there are the subtle changes in regional floodplain slopes caused by the flexing of the Indian tectonic plate in response to the growth and erosion of the adjacent mountain belt, whose great mass acts downwards — like a swimmer on the end of a diving board. The longitudinal sediment-filled troughs produced by such flexural downbending⁹, known as foreland basins, are particularly prone to major diversion of river systems flowing along the axis of the basin. Unlike steep rivers in tectonically uplifting mountain areas, which cut down into bedrock like cheese wires, low-gradient rivers in foreland basins are easily deflected. Continental-scale diversion might also result from the effects of climate change on river discharge, which may allow one river

to dominate its neighbours and capture their drainage systems.

Clearly, investigators attempting to interpret the sedimentary record of the deep sea must be careful to disentangle the effects of climate change, variations in tectonically driven erosion, and continental-scale switches of river drainage. The use of a range of isotopic signatures in river and deep-sea sediments will help in this challenging undertaking.

Philip A. Allen is in the Department of Earth Science and Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
e-mail: philip.allen@imperial.ac.uk

1. Clift, P. D. & Blusztajn, J. *Nature* **438**, 1001–1003 (2005).
2. Milliman, J. D. & Meade, R. H. *J. Geol.* **91**, 1–21 (1983).
3. Milliman, J. D. & Syvitski, J. P. M. *J. Geol.* **100**, 525–544 (1992).
4. Walling, D. E. & Webb, B. W. in *Erosion and Sediment Yield: Global and Regional Perspectives* (eds Walling, D. E. & Webb, B. W.) 3–19 (Int. Assoc. Hydrol. Sci., 1996).
5. Métivier, F., Gaudemer, Y., Tapponier, P. & Klein, M. *Geophys. J. Int.* **137**, 280–318 (1999).
6. DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 249–252 (1976).
7. Beck, R. A. & Burbank, D. W. *Geol. Soc. Am. Abstr. with Prog.* **22**, A238 (1990).
8. Reynolds, R. G. H. *Geol. Bull. Univ. Peshawar* **14**, 141–150 (1981).
9. Burbank, D. W., Beck, R. A. & Mulder, T. in *The Tectonic Evolution of Asia* (eds Yin, A. & Harrison, T. A.) 149–188 (Cambridge Univ. Press, 1996).

DEVELOPMENTAL BIOLOGY

A message to the back side

Wolfgang Driever

Vertebrate embryos from fish to mammals seem to use different routes to work out which way is up and which side is front. Yet a novel system involved in defining the dorsal side of fish might be conserved in mammals.

Vertebrates have many developmental processes in common; but so far, no unifying mechanism that specifies the dorsal–ventral (back-to-belly) axis in the vertebrate early embryo has been found. Egg cells, or oocytes, are in general roughly spherical and have only one axis: animal–vegetal, often characterized by the cell nucleus being in the ‘animal’ portion and away from the ‘vegetal’ yolk-rich pole. In amphibians and fish, after fertilization certain protein signals are physically transported from the vegetal region to the future dorsal side, contributing to the specification of dorsal. By contrast, the mechanisms of axis formation in mammals are not understood.

On page 1030 of this issue, Gore and colleagues¹ present evidence from zebrafish that *nodal* messenger RNAs, which encode the dorsal signal protein Nodal, are progressively localized to the cells that go on to form the dorsal side. Surprisingly, this dorsal localization also occurs if sequence elements from human *nodal* mRNA are used instead of the zebrafish ones, indicating an evolutionarily conserved mechanism.

The zebrafish version of Nodal is formally called Ndr1 (for Nodal-related 1) and, like the Nodal proteins found in all other vertebrates investigated, it is involved in specifying dorsal structures and two of the major embryonic

tissues, mesoderm and endoderm. This Nodal activity occurs at stages when the embryo has more than 1,000 cells and expresses its own (zygotic) genes^{2,3}, as opposed to relying on mRNA generated during oogenesis (that is, maternal mRNA). Gore and Sampath⁴ previously showed that the distribution of maternal *ndr1* is uniform throughout oogenesis and up to fertilization. In the current work, Gore *et al.*¹ found that during the first three cell divisions of the zygote, maternal *ndr1* mRNA is localized to two of the eight cells of the embryo — the two cells that will later form the dorsal side of the embryo (Fig. 1). Furthermore, they injected fluorescent *ndr1* mRNA into the one-cell embryos and showed that its dorsal movement over the next cell cycles depends on the microtubules of the cytoskeleton, the cell's internal transport tracks.

Gore *et al.* demonstrated that maternal *ndr1* mRNA is involved in dorsal specification, because inhibiting the synthesis of protein from the maternal as well as the zygotic *ndr1* mRNAs leads to the formation of radially symmetric embryos that lack all dorsal structures. By contrast, in embryos in which just the zygotic *ndr1* had been mutated only some of the dorsal structures were lost. The authors have not analysed the fates of embryos mutant for both the maternal and zygotic *ndr1* contribution, which would provide the ultimate proof for an axis-specifying function of Ndr1.

The Wnt signalling pathway is the best-characterized pathway involved in dorsal specification. Activation of one of its components, β -catenin, is both necessary and sufficient to induce dorsal structures in vertebrates ranging from fish to birds⁵. To see whether β -catenin is involved in *ndr1* mRNA localization, Gore *et al.*¹ examined zebrafish with a mutation that eliminates maternal β -catenin from oogenesis onwards. They showed that labelled *ndr1* mRNA still moves to the same two cells, so this mechanism is independent of the Wnt pathway during dorsal specification. However, studies of the mutant also showed that during later development, β -catenin contributes to the expression of zygotic *ndr1* at the dorsal side of the embryo, which will make it difficult to separate out the two pathways. It is unclear what the targets of the localized maternal *ndr1* message may be. Nodal can induce its own expression⁶, but maternal Nodal seems not to be sufficient to induce zygotic *ndr1* expression in the absence of β -catenin activity⁷. Perhaps Nodal signalling derived from maternal mRNA in dorsal cells, even though at low levels, contributes to a propensity to respond fully to other signalling pathways.

This is the first time that dynamic mRNA localization has been reported to be involved in specification of the vertebrate dorsal-ventral axis. Many invertebrates use mRNA localization to define dorsal-ventral or anterior-posterior axes — and are very creative with regard to mechanism. For example, embryos of the fruitfly *Drosophila*⁸ use active mRNA

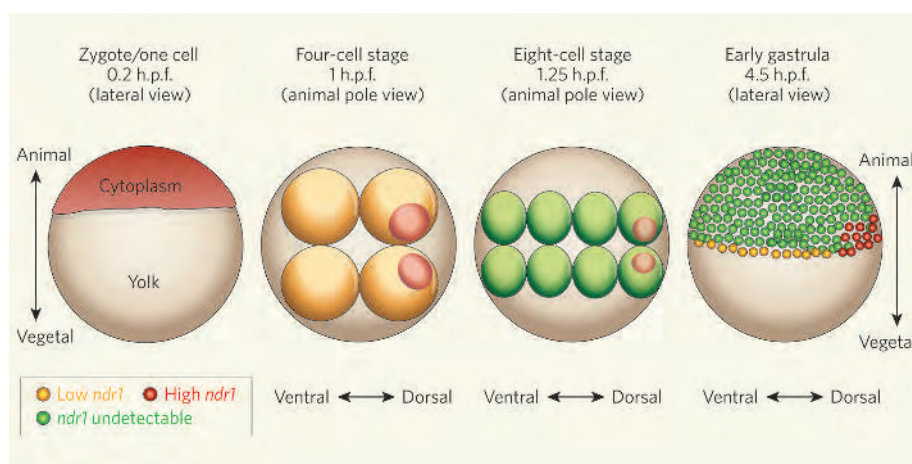


Figure 1 | Maternal mRNA that encodes Nodal moves to the dorsal side. Gore *et al.*¹ followed the progress of maternal as well as injected fluorescent *ndr1* messenger RNA through the first few cell divisions of the zebrafish embryo. At the eight-cell stage the mRNA is found only in the cells that will go on to form the dorsal side of the fish. (h.p.f., hours post-fertilization.)

transport involving local attachment to the cytoskeleton (the *bicoid* mRNA), as well as local mRNA synthesis after moving a nucleus (*gurken* mRNA) and local stabilization of mRNA (*nanos* mRNA).

Gore *et al.*¹ next examined a non-coding portion (the 3' region) of the *nodal* mRNAs from eight vertebrate species, from fish to humans, because such regions mediate sub-cellular mRNA localization in other systems⁸. Three short elements in this region are evolutionarily conserved among the species. The authors showed that the *ndr1* 3' elements are required for dorsal localization, and that both zebrafish and human 3' elements can direct an mRNA to the dorsal side of zebrafish embryos.

Does this result mean that similar mechanisms of dorsal specification act in fish and mammals? If so, *nodal* mRNA distribution would have to be asymmetrical in the mouse zygote at the one-cell stage, and features of asymmetry must be expressed from the two-cell stage onwards. Such early asymmetry is fiercely debated, with several developmental biologists reporting early axis specification⁹, whereas others conclude that the earliest polarity is only established three days later, once the cells have formed the blastocyst (32 cells and greater)¹⁰. Although *nodal* mRNA has not been detected at the blastocyst or earlier stages, it is present in embryonic stem cells and shortly after implantation¹¹. Interestingly, there is a hypothesis that, at the level of developmental mechanisms, exactly this stage would be conceptually equivalent to the eggs of most non-mammalian species¹⁰.

One can only speculate on how mechanisms localizing *nodal* mRNA within cells of a mammalian implantation-stage embryo might contribute to dorsal specification in a cellular context — proper cellular localization may control mRNA stability and efficiency of expression as well as transport of Nodal. Alternatively, Nodal at the zygote stage (previously undetected) and its asymmetric distribution

may propagate through the positive and negative autoregulatory loops intrinsic to the Nodal signalling pathway, and slowly establish a bias between dorsal and ventral. Such an initially weak dorsal bias has been postulated to explain the regulatory nature of the early mammalian embryo¹². Zygote-stage asymmetries in Nodal distribution might also contribute to asymmetries in intermediary targets rather than continuously maintaining Nodal expression. Such intermediaries could then contribute to dorsal specification by novel mechanisms.

The movement of *ndr1* mRNA in zebrafish zygotes is an exciting lead into potential unifying mechanisms of axis formation in vertebrate embryology. However, mechanisms of *nodal* mRNA localization need to be better understood, and demonstrated in mammals. Detecting Nodal protein distribution is a notoriously difficult task, but it must be done, or some reliable reporter found that can monitor slight differences in Nodal signalling. Finally, identifying targets of Nodal's early axis-determining activity could reveal whether there truly are conserved pathways of axis specification in vertebrates. ■

Wolfgang Driever is in the Developmental Biology Unit, Department of Biology, University of Freiburg, Hauptstrasse 1, D-79104 Freiburg, Germany.
e-mail: driever@biologie.uni-freiburg.de

- Gore, A. V. *et al.* *Nature* **438**, 1030–1035 (2005).
- Feldman, B. *et al.* *Nature* **395**, 181–185 (1998).
- Chen, Y. & Schier, A. F. *Nature* **411**, 607–610 (2001).
- Gore, A. V. & Sampath, K. *Mech. Dev.* **112**, 153–156 (2002).
- DeRobertis, E. M., Jarrain, J., Oelgeschläger, M. & Wessely, O. *Nature Rev. Genet.* **1**, 171–181 (2000).
- Pogoda, H.-M., Solnica-Krezel, L., Driever, W. & Meyer, D. *Curr. Biol.* **10**, 1041–1049 (2000).
- Kelly, C. *et al.* *Development* **127**, 3899–3911 (2000).
- St Johnston, D. *Nature Rev. Mol. Cell Biol.* **6**, 363–375 (2005).
- Plusa, B. *et al.* *Nature* **434**, 391–395 (2005).
- Motosugi, N. *et al.* *Genes Dev.* **19**, 1081–1092 (2005).
- Varlet, I., Collignon, J. & Robertson, E. J. *Development* **124**, 1033–1044 (1997).
- Zernicka-Goetz, M. *Development* **129**, 815–829 (2002).

OBITUARY

Richard Southwood (1931–2005)

Entomologist, ecologist and science policy adviser.

Modern ecology emerged in the 1960s and 1970s as a fusion of scientific natural history, applied biology and more rigorous approaches from population biology and mathematics. Richard (Dick) Southwood, who died on 26 October, was a major figure in forging the new subject. He was influential both for his own research and for the research groups that he built and fostered.

Southwood's background was as an entomologist, in particular as an expert on the Heteroptera or true bugs, a group that contains many pest species. His PhD on their systematics and ecology, undertaken at Rothamsted Experimental Station in Hertfordshire, UK, was very much in the mould of classical entomology; likewise the first research he did when, in 1955, he moved to Imperial College London to study the cereal pest *Oscinella*, the frit (not fruit) fly.

Although he continued to study applied problems, for example mosquito dynamics and improving the habitat for partridges, Southwood increasingly turned to the more fundamental issues in ecology for which he is best known. Here his work was invariably motivated by his encyclopaedic knowledge of insect faunas. In the early 1960s he began studying why different species of tree support remarkably different numbers of herbivorous insect species. Using many sources, he compiled a database for insects on trees in different regions, which indicated the roles of history, biochemistry and whether a tree species has few or many close relatives in Britain. Today, this method would be called macroecology, but then it was an unusual and new approach.

Over the next 20 years, studies of insect herbivore communities by Southwood and the many others he inspired became test cases in community ecology. They were used, for example, to explore the degree to which the theories of island biogeography could be applied to non-overlapping resource types and the extent to which terrestrial insect communities are structured by interactions with their environment, as opposed to being mere assemblages of species drawn at random from those that can survive in a particular habitat.

Heteropteran bugs vary greatly in their dispersal capabilities, with many being strong fliers and others completely wingless. Ever since his PhD, Southwood had been interested in the evolution of life histories, and in particular how the scale and spatial structure of the habitat of different species led to selection for varying reproductive and

dispersal strategies. He summarized this approach in 1977, in a major review article, "Habitat, the templet for ecological strategies?", which became a citation classic. This work was notable for linking the traditional informal approach to studying life histories with the growing mathematical theory based on optimality arguments and trade-offs between different demands on an organism's limiting resources.

Southwood remained at Imperial until 1979, building a strong group in pure and applied ecology. For the rest of his career he was based at the University of Oxford, initially as head of the Department of Zoology, where he continued to apply his talents as a scientific leader and group builder. He eventually became vice-chancellor of the university, a post that, particularly at Oxford, requires the wisdom of Solomon. He also became increasingly involved in scientific policy, and was an adviser to the UK government on many issues involving the environment and health. In the early 1980s, he chaired the Royal Commission on Environmental Pollution that produced a report on the consequences of lead in petrol. It was this report that led to the adoption of lead-free petrol in the United Kingdom, as well as in many other countries.

In 1988, Southwood was asked to chair the working group set up to advise on control measures and the risks of bovine spongiform encephalopathy (BSE), which was ravaging UK cattle herds. This was a highly sensitive issue, with obvious risks for public health if the disease could jump the species barrier to humans, but also with serious economic consequences for British farming if the risks were overstated. Southwood's group strongly criticized the changes in rendering practices that had led to cattle and sheep products being fed to cattle, and recommended that certain parts of the cow should not enter the human food chain. Most of these recommendations were implemented, although with less urgency than Southwood wanted; he was particularly critical of the decision to compensate farmers for only half of the economic value of infected cows, a clear disincentive for farmers to report the disease.

The working group also considered that the risk of BSE jumping to humans was most unlikely, although it stressed the importance of reducing that risk further. This opinion was shared by the vast majority of scientists at the time, and was particularly influenced by the fact that scrapie in sheep, a disease with a



similar aetiology to BSE, had never been a problem for human health. To Southwood's dismay, his report was used by government without equivocation to persuade people of the near impossibility of human infection. When human infections did occur, and when it was unclear how many people would become infected, Southwood's group was criticized for not warning more strongly of the possible dangers. The episode highlights the great difficulties in providing scientific advice in the face of uncertainty, and the dangers of nuanced scientific arguments being reinterpreted in the political arena.

Dick Southwood was a man of immense charm who will be remembered with great fondness, and some awe, by generations of ecologists. He never lost the love of natural history that filled his summer holidays as a schoolboy in Kent, and which informed so much of his scientific research. The handbook for the identification of British Heteroptera that he wrote in 1959 with Dennis Leston, *Land and Water Bugs of the British Isles*, is still the standard work on the group.

He was also an inspired teacher, and wrote an influential textbook on ecological methods. Students found his exuberance for science and natural history infectious. Mischievously, when teaching sampling techniques on Imperial College field courses, he would contrive that 'randomly' placed quadrats would include the most interesting plants or insects in the site — to this day, we teach our students the difference between randomized, stratified and southwoodian sampling!

Charles Godfray and Michael Hassell

Charles Godfray and Michael Hassell are in the Division of Biology, Silwood Park Campus, Imperial College London, Ascot SL5 7PY, UK. e-mails: c.godfray@imperial.ac.uk; m.hassell@imperial.ac.uk

UPPA.CO.UK

BRIEF COMMUNICATIONS

Post-spawning egg care by a squid

Spying on a brooding deep-sea squid reveals that it cradles and aerates its eggs while they mature.

Gonatus onyx is one of the most abundant cephalopods in the Pacific and Atlantic Oceans¹ and is an important prey species for a variety of vertebrate predators^{2,3}, but a full understanding of its life history has been hampered because spawning occurs at great depths^{4,5}, where observation is difficult. Here we describe post-spawning egg care, or brooding, in this deep-sea squid. Our finding is unexpected because this behaviour differs from the reproductive habits of all other known squid species.

It was initially assumed that gonatids, like other squids, deposited their eggs on the sea floor and left them to develop on their own¹. A pelagic egg-brooding habit had been proposed for gonatid squids⁶, but could not be verified in the absence of direct observation. It had also been questioned because some aspects of the squids' biology seem to preclude brooding⁷ — principally, the degeneration of musculature following sexual maturation was presumed to limit locomotion and render squids unfit for egg protection^{5,7}.

However, we observed five squids, each holding an egg mass in its arms, at depths between 1,539 and 2,522 m in Monterey Canyon, off California, accessed with the ROV *Tiburon*⁸. Eggs or hatchlings and two adults were collected. The squids used hooks on their arms to hold the egg mass, which consists of two thin membranes in a continuous flat sheet, open at both the distal and proximal ends. The egg mass forms a hollow tube that extends from the mouth to well beyond the end of the arms and contains about 2,000–3,000 eggs (Fig. 1a, b).

Repeated extension of the arms (at intervals of about 30 to 40 s) flushed water through the egg mass: this behaviour probably served to aerate the eggs in the hypoxic midwaters found off California^{9,10}. Aggressive arm movements and escape swimming caused partial disintegration of the more mature egg masses and hatching of the released eggs (Fig. 1c–e). (For movies, see supplementary information.)

The escape responses seen here, considered with the apparent stage of egg development,

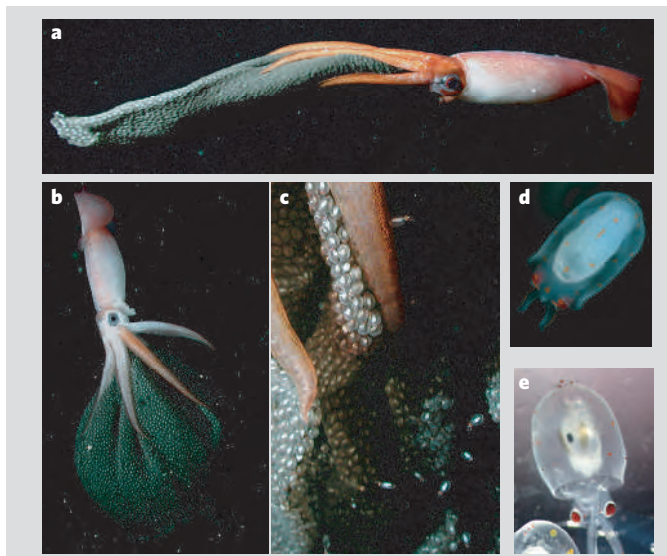


Figure 1 | Egg-brooding adult female *Gonatus onyx* (Cephalopoda: Teuthoidea), photographed *in situ*. Mantle length is about 145 mm. **a**, The squid in a horizontal resting position at 2,522 m depth. **b**, Squid holding a tubular egg mass, and **c**, hatchlings being released at 1,539 m depth. **d**, **e**, Hatched embryos (about 3 mm in length) at **d**, an intermediate stage of development, and **e**, an advanced stage of development. The *in situ* temperature (1.7–3.0 °C) and oxygen concentration (45–90 $\mu\text{mol l}^{-1}$) were measured using a SeaBird SBE-9 conductivity–temperature–depth unit with an oxygen sensor. Additional visual observations (see movie in supplementary information) were recorded with a high-resolution, three-chip video camera transmitting to the RV *Western Flyer* by fibre-optic cable. Assisted by vessel crew and ROV pilots.

indicate that the squid may have undergone a gradual degeneration of locomotory capacity. For example, a specimen bearing undeveloped eggs made a vigorous escape by using fin and mantle contractions, whereas those with advanced embryos (Fig. 1e) showed only respiratory mantle contractions and did not move away; another squid with eggs at an intermediate stage of development retained some locomotory ability (Fig. 1d) (for movies, see supplementary information). Activities of metabolic enzymes in locomotory muscles of spawned female *G. onyx* are lower in specimens that have more advanced embryos⁵.

Low temperature and large eggs will prolong development in *G. onyx*^{7,11}. The abundance of juveniles in near-surface waters peaks seasonally from April through to July¹, which may indicate a yearly cycle with an egg-development period of 6 to 9 months^{5,6}. This estimate is roughly consistent with the timing of our observations. Gonatid squids have sufficient lipid stores to fuel metabolism for a long brooding period, and the gradual decline

in mantle muscle and digestive-gland condition indicates a progressive use of energy stores over a long egg-brooding period⁶. However, two specimens captured within days of each other carried embryos at different stages, so development is apparently not well synchronized.

We have demonstrated active post-spawning egg care in *G. onyx* and this is, to our knowledge, the first case reported in squids. The value of this strategy in the deep sea is supported by strong convergence with distantly related taxa¹² and we expect it to be found in other squids. Despite retaining some capacity for swimming, the relatively immobile brooding squids are found within the usual diving range of whales and elephant seals and so may provide an easy target for such 'mesopelagic' mammals¹³. Gonatid squids and other ontogenetic migrators therefore represent a direct energetic, as well as trophic, link between deep and shallow biomes.

Brad A. Seibel*†, **Bruce H. Robison***, **Steven H. D. Haddock***

*Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA
e-mail: seibel@uri.edu

†Present address: Biological Sciences Department, University of Rhode Island, Kingston, Rhode Island 02881, USA

- Okutani, T., Kubodera, T. & Jefferts, K. *Bull. Ocean. Res. Inst. Univ. Tokyo* **26**, 159–192 (1988).
- Clarke, M. R. *Phil. Trans. R. Soc. Lond. B* **351**, 979–983 (1996).
- Hooker, S. K., Whitehead, H. & Gowans, S. *Biol. Conserv.* **104**, 51–48 (2002).
- Arkhipkin, A. I. & Bjorke, H. *Polar Biol.* **22**, 347–365 (1999).
- Hunt, J. C. & Seibel, B. A. *Mar. Biol.* **136**, 543–552 (2000).
- Seibel, B. A., Hochberg, F. G. & Carlini, D. B. *Mar. Biol.* **137**, 519–526 (2000).
- Nesis, K. N. *Russ. J. Mar. Biol.* **25**, 499–506 (1999).
- Robison, B. H. *Mar. Technol. Soc. J.* **26**, 32–39 (1993).
- Strathmann, R. R. & Strathmann, M. F. *J. Mar. Biol. Assoc. UK* **75**, 413–428 (1995).
- Childress, J. J. & Seibel, B. A. *J. Exp. Biol.* **201**, 1223–1232 (1998).
- Boletzky, S. V. *Antarctic Sci.* **6**, 139–142 (1994).
- Childress, J. J. & Price, M. H. *Mar. Biol.* **76**, 165–177 (1983).
- Hochachka, P. W. *Experientia* **48**, 570–574 (1992).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/438929a

COLLOID SCIENCE

Non-spherical bubbles

Surface tension gives gas bubbles their perfect spherical shape by minimizing the surface area for a given volume¹. Here we show that gas bubbles and liquid drops can exist in stable, non-spherical shapes if the surface is covered, or 'armoured', with a close-packed monolayer of particles. When two spherical armoured bubbles are fused, jamming of the particles on the interface supports the unequal stresses that are necessary to stabilize a non-spherical shape.

We have previously described a microfluidic method for producing spherical armoured bubbles that are all the same size². The rigid particles straddle the gas-liquid interface and have mechanical properties distinct from either constituent, forming what we call an interfacial composite material.

We find that fusion of these armoured bubbles, achieved by squeezing the bubbles between two glass plates, produces a stable ellipsoidal shape (Fig. 1 a–c) (for methods, see supplementary information). The fused armoured bubble is unable to relax to a spherical shape by expelling particles: instead, the jamming³ of the particles on the closed interface, which is mediated by surface tension, leads to non-minimal shapes.

The non-trivial geometry of these bubbles provides a natural means of understanding

the state of stress in the interfacial composite material. A balance of normal stresses at the bubble surface demands that

$$\Delta P = \frac{\sigma_1}{R_1} + \frac{\sigma_2}{R_2}$$

where ΔP is the pressure jump across the surface, R_1 and R_2 are the local principal radii of curvature, and σ_1 and σ_2 are the corresponding principal resultants of surface stress. Therefore, if $R_1 \neq R_2$, as is the case for non-spherical bubbles, then $\sigma_1 \neq \sigma_2$. A simple fluid interface at equilibrium cannot support unequal stresses³. But the bubble does, because of steric jamming⁴ of the armour particles, so we term the interfacial composite material a solid.

The armoured bubbles can be remodelled into various stable anisotropic shapes because the interfacial composite material is able to undergo extensive particle-scale rearrangements in order to accommodate external inhomogeneous stresses (our manuscript in preparation). These shape changes occur with apparently no hysteresis and at relatively low forces, which is equivalent to perfect plasticity in continuum mechanics.

High aspect-ratio shapes with saddle curvature can be maintained on the armoured

bubbles (Fig. 1d). This feature may be exploited to change the topology of the bubble by introducing a hole into the object, thereby creating a stable genus-1 toroid (Fig. 1e). The change in topology is irreversible⁵, and seems to be the only permanent change associated with the manipulation of the interfacial composite material.

We have found that interfacial jamming is a general phenomenon that occurs with particle types such as polymethylmethacrylate, gold and zirconium oxide, and that spans four orders of magnitude in particle and bubble sizes. Similar effects are evident with liquid droplets of mineral oil that are covered with rigid particles (Fig. 1f).

Stable, non-spherical shapes of pressurized systems that have no obvious source of a stress-bearing network have been reported for dirty air bubbles in the ocean⁶ and for various cellular organelles⁷. Also, systems such as gelled lipids on air bubbles⁸ and protein-coated vesicles⁹ show plasticity. We propose that a generic interfacial jamming transition may explain the mechanical properties and structural stability of these diverse systems.

Anand Bala Subramaniam, Manouk Abkarian, L. Mahadevan, Howard A. Stone

Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

e-mail: has@deas.harvard.edu

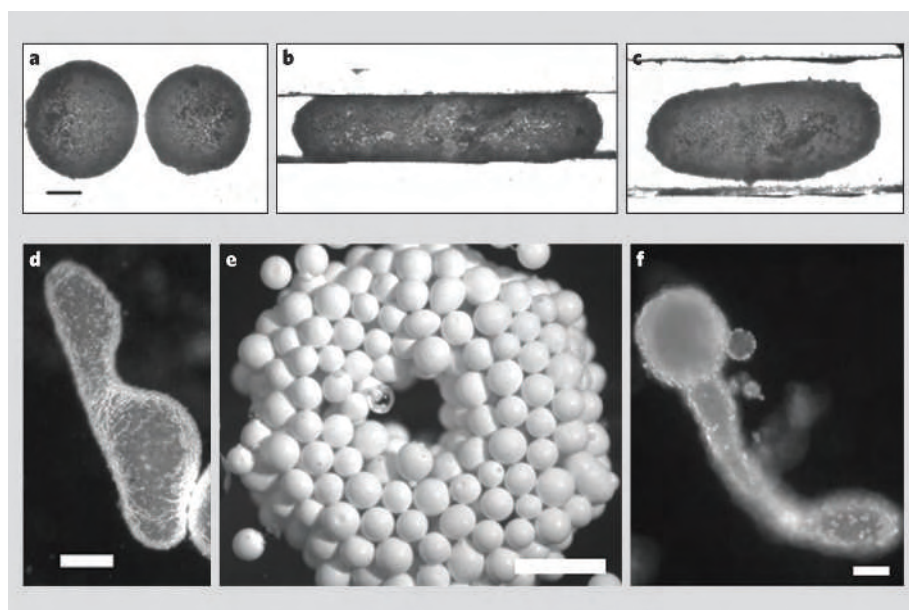


Figure 1 | Non-spherical gas bubbles. In a–d, the bubbles are covered with charge-stabilized, fluorescent polystyrene beads, each of 2.6 μm diameter. **a**, Two initially spherical armoured bubbles. **b**, The bubbles are compressed between two glass plates (see supplementary information for details), which exposes naked interfaces that spontaneously coalesce. **c**, The gas bubble maintains a stable ellipsoidal shape even after the side plates are removed. **d**, Armoured bubble with a stable saddle shape. **e**, The ability to maintain a saddle curvature allows a hole to be introduced into the bubble to create a permanent change of topology into a genus-1 toroid; here the particles are ground zirconium, of average diameter 200 μm. **f**, Non-spherical shapes can be similarly maintained on mineral-oil droplets in water armoured with 4.0-μm fluorescent polystyrene particles. Scale bars (μm): a–c, 100; d, 200; e, 500; and f, 16.

- Boys, C. V. *Soap Bubbles — Their Colours and the Forces which Mould Them* (Dover, New York, 1959).
- Bala Subramaniam, A., Abkarian, M. & Stone, H. A. *Nature Mater.* **4**, 553–556 (2005).
- Vella, D., Aussillous, P. & Mahadevan, L. *Europhys. Lett.* **68**, 212–218 (2004).
- Liu, A. J. & Nagel, S. R. *Nature* **396**, 21–22 (1998).
- Alexandrov, P. S. *Elementary Concepts of Topology* (Dover, New York, 1961).
- Johnson, B. D. & Cooke, R. C. *Science* **213**, 209–211 (1981).
- Joachim, S., Jokitalo, E., Pypaert, M. & Warren, G. *Nature* **407**, 1022–1026 (2000).
- Kim, D. H., Costello, M. J., Duncan, P. B. & Needham, D. *Langmuir* **19**, 8455–8466 (2003).
- Ratanabangkoon, P., Gropper, M., Merkel, R., Sackmann, E. & Gast, A. P. *Langmuir* **19**, 1054–1062 (2003).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/438930a

ERRATUM

Nanoscale hydrodynamics: Enhanced flow in carbon nanotubes

Mainak Majumder, Nitin Chopra, Rodney Andrews, Bruce J. Hinds
Nature **438**, 44 (2005)

In Table 1, slip lengths are in micrometres (and not millimetres, as published).

doi:10.1038/438930b

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

SEISMOLOGY

Tectonic strain in plate interiors?

Arising from: R. Smalley Jr, M. A. Ellis, J. Paul & R. B. Van Arsdale *Nature* 435, 1088–1090 (2005)

It is not fully understood how or why the inner areas of tectonic plates deform, leading to large, although infrequent, earthquakes. Smalley *et al.*¹ offer a potential breakthrough by suggesting that surface deformation in the central United States accumulates at rates comparable to those across plate boundaries. However, we find no statistically significant deformation in three independent analyses of the data set used by Smalley *et al.*, and conclude therefore that only the upper bounds of magnitude and repeat time for large earthquakes can be inferred at present.

The occurrence of earthquakes at the interior of tectonic plates — assumed to be rigid in conventional plate tectonic theory — indicates that stresses within plates accumulate on faults and are released during large, but rare, events. How this cycle relates to the slow deformation of plate interiors is unknown, posing significant difficulties for understanding the associated hazards. Stakes are high because several, now densely populated, intraplate areas have been struck in the past by large earthquakes, including in the central United States in 1811–12, in Basel, Switzerland, in 1356, and in Newcastle, Australia, in 1989. Geophysicists are now using the global posi-

tioning system (GPS) to quantify strain in plate interiors in the hope of relating it to stress build-up on seismogenic faults.

Smalley *et al.* report significant strain from GPS measurements in the New Madrid seismic zone (NMSZ) of the central United States. They interpret their findings as indicating deformation rates comparable to those observed at much more seismically active plate boundaries¹. If confirmed, this result could give insight into the processes that drive the occurrence of large earthquakes in plate interiors, and provide new quantitative information for seismic-hazard estimation in the New Madrid area¹.

However, independent analyses of the same data, performed by three independent groups using different analysis software and processing strategies, reveal no statistically significant site motions or strains (Fig. 1), with an average weighted misfit to a rigid-plate behaviour of 1.4 mm yr^{-1} (95% confidence). In particular, the shortening between sites RLAP and NWCC, used by Smalley *et al.*¹ as their primary argument for strain accumulation on the Reelfoot fault, is of marginal significance ($1.7 \pm 2.0 \text{ mm yr}^{-1}$; 95% confidence) and largely reflects an unexplained offset that

occurred between mid-2001 and early 2002 (Fig. 1, inset). The same analyses, using 156 GPS sites distributed throughout the central and eastern United States, find no spatially coherent deviation from rigid behaviour in the far field of the NMSZ either, apart from effects due to the removal of glacial loads, with an average weighted misfit to a rigid-plate model of 1.4 mm yr^{-1} (95% confidence) as well (further details are available from the authors).

Detecting motion depends critically on the assumed uncertainties of site velocities, which decrease as data span longer times. Hence the present data do not preclude the possibility that a statistically significant tectonic signal may emerge in the future. We shall then face the challenge of deciding whether the deformation represents strain accumulating for release in a future earthquake¹ or long-term relaxation after the 1811–12 earthquakes^{2,3}.

Is an upper bound of 1.4 mm yr^{-1} of motion across the NMSZ consistent with longer-term data from palaeo-earthquakes in the central United States?¹ Assuming that characteristic earthquakes repeat regularly in the NMSZ (probably an oversimplification, although it is one used in National Earthquake Hazard maps), this leads to a minimum repeat time of about 600–1,500 years, consistent with earlier estimates⁴ based on the palaeoseismic history⁵ if one assumes occurrence of earthquakes of magnitude 7, with 1–2 m of co-seismic slip⁴.

Although intraplate earthquakes indicate that tectonic stresses within plate interiors accumulate on faults and are released during large, infrequent events, deviations from rigid behaviour in the central United States and several other major plates^{6,7} are below the current resolution of GPS measurements and do not reflect this cycle — at least not on a timescale of a decade or less. Longer observation spans and further improvement of geodetic techniques are needed to understand where, why and how much strain concentrates in plate interiors.

E. Calais*, G. Mattioli†, C. DeMets‡, J.-M. Nocquet§, S. Stein||, A. Newman¶, P. Rydelek#

*Department of Earth and Atmospheric Sciences, Purdue University, West Lafayette, Indiana 47907, USA

e-mail: ecalais@purdue.edu

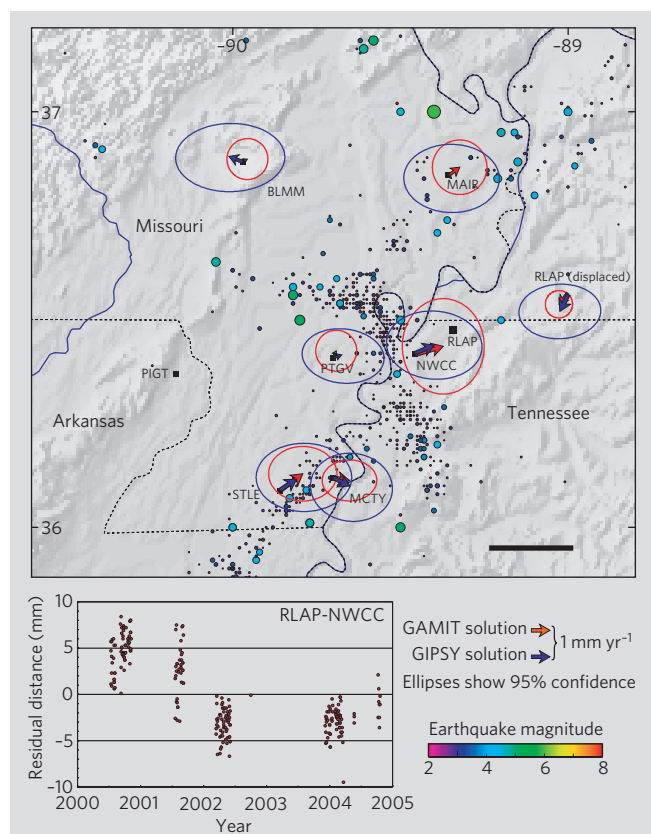
†Department of Geology, University of Arkansas, Fayetteville, Arkansas 72701, USA

‡Department of Geology and Geophysics, University of Wisconsin-Madison, Wisconsin 53706, USA

§CNRS, Geosciences Azur, 06560 Valbonne, France

||Department of Geological Sciences,

Figure 1 | Velocities and associated uncertainties (95% confidence) at continuous GPS sites in the New Madrid seismic zone. To perform these analyses, two different software packages (GAMIT and GIPSY) were used. Site velocities are within their error ellipses and hence show no statistically significant motion. Filled coloured circles show regional seismicity (United States Geological Survey catalogues; details of site names are listed in Table 1 of ref. 1, except BLMM). The different arrow types represent two independent solutions. Scale bar, 20 km. Inset, time series of daily baseline length estimates between sites RLAP and NWCC after removal of a mean. Error bars on daily estimates, omitted for the sake of clarity, are of the order of 2–3 mm.



Northwestern University, Evanston,
Illinois 0208-2150, USA
¶School of Earth and Atmospheric Sciences,
Georgia Institute of Technology, Atlanta,
Georgia 30332, USA
#Center for Earthquake Research, University of
Memphis, Memphis, Tennessee 38152, USA

1. Smalley, R. Jr, Ellis, M. A., Paul, J. & Van Arsdale, R. B. *Nature* **435**, 1088–1090 (2005).

2. Rydelek, P. & Pollitz, F. F. *Geophys. Res. Lett.* **21**, 2302–2306 (1994).
3. Kenner, S. J. & Segall, P. *Science* **289**, 2329–2332 (2000).
4. Newman, A. et al. *Science* **284**, 619–621 (1999).
5. Tuttle, M. P. & Schweig, E. S. *Geology* **23**, 253–256 (1995).
6. Nocquet, J.-M., Calais, E. & Parsons, B. *Geophys. Res. Lett.* **32**, doi:10.1029/2004GL022174 (2005).
7. Beavan, J., Tregoning, P., Bevis, M., Kato, T. & Meertens, C. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000282 (2002).

doi:10.1038/nature04428

SEISMOLOGY

Smalley et al. reply

Replying to: E. Calais et al. *Nature* **438**, doi:10.1038/nature04428 (2005)

The independent analyses of GAMA (global positioning system (GPS) array in mid-America) data by Calais et al.¹ demonstrate the difficulties in determining patterns of rational deformation within otherwise rigid plates. We are a long way from incorporating this type of information into seismic-hazard analysis, and we agree that longer time spans and improved spatial coverage with geodetic-quality data are required in order to gather the observations necessary to start modelling and understanding this enigmatic region.

The uncertainties in the analyses of Calais et al.¹ and in our own analysis² are reported at the 1-sigma level, but are shown at 95% confidence level on the maps (Fig. 2 of ref. 2 mistakenly identifies the uncertainties displayed on the map as 1-sigma rather than as their correct 95% confidence interval). There is no disagreement between the two sets of results^{1,2} for the far-field component of the array, where uncertainties in both are larger than surface velocities. The differences arise between analyses in the critical near-field sites, which straddle

the active faults. Velocity vectors and errors at these sites are remarkably close for the two GAMIT solutions: differences arise from the slightly larger uncertainty in the results of Calais et al.¹

Uncertainties in GPS analyses are poorly understood (see the differences between GIPSY and GAMIT analyses reported in refs 1, 2) and depend on the size of the array being considered and the pattern of deformation being sought. We illustrate this with a simple thought experiment: consider a least-squares fit to a straight line using a set of 150 points with a given error distribution, then add to this a second set of 10 points that span a limited range and for which the slope differs by 10%; it will be almost impossible, by statistical means, to distinguish the second set in the combined set. The statistics of the larger set will dominate the uncertainties of the smaller, and the only way to distinguish the two sets is to limit the data to reveal (perhaps serendipitously) the smaller and significant data set. This effect will be compounded if, in

an analysis of a GPS network, the station spacing is larger than the scale expected of local deformation, so that the large-array analysis will probably be aliased.

The illustrated recurrence interval¹, based on an assumed upper bound for fault slip of 1.4 mm yr⁻¹, is limited by the assumption that strain accumulation is linear over time (processes of this sort can be nonlinear), and by palaeoseismological evidence indicating an average recurrence (albeit limited by sparse data) of about 500 years (not 600–1,500 years¹) over the past 2,000 years². Such recurrence would, simplistically, require so-called fault-slip rates greater than 4 mm yr⁻¹. However, debating these few data in terms of a specific seismic hazard is risky (and we avoided it earlier²) because the source of such displacements is unknown²: they are snapshots of a potentially complex spatial and temporal pattern of fault-related displacements.

The relationship of our derived displacements and the well known active faults in the New Madrid region remain a compelling argument to us that the system is active, a conclusion borne out by a decade of geological results in the region². Neither we nor anyone else can so far explain this apparent local deformation — in the spirit of Galileo, “and yet it moves”.

R. Smalley Jr*†, M. A. Ellis*†, J. Paul*,
R. B. Van Arsdale†

*Center of Earthquake Research and Information,
†Department of Earth Sciences, The University of
Memphis, Memphis, Tennessee 38152, USA
e-mail: mellis@nsf.gov

1. Calais, E. et al. *Nature* **438**, doi:10.1038/nature04428 (2005).
2. Smalley, R. Jr, Ellis, M. A., Paul, J. & Van Arsdale R. B. *Nature* **435**, 1088–1090 (2005).

doi:10.1038/nature04429

**Cover illustration**

Developing mouse retinal blood vessels grow towards areas of low oxygen concentration. (Courtesy of M. Fruttiger, Kings College London/Development.)

Editor, Nature

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Lesley Anson

Consultant Editor

Ritu Dhand

Production Editor

Maria Hodges

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Layouts

Nadia Rooney

Sponsorship

Claire Hines
Claudia Banks

Production

Sue Gray

Marketing

Robin Brown

Editorial Assistant

Laura Shaw

ANGIOGENESIS

Blood vessels are a complex network of tubes that carry oxygenated blood and nutrients throughout our bodies. If laid end to end, the vessels from a typical adult would circle the Earth twice. It comes as no surprise, then, that the process of growing new blood vessels — angiogenesis — is a fundamental biological mechanism that results in serious disease when it goes awry. Indeed, more than US\$4 billion has been invested in the research and development of medicines to promote or reduce angiogenesis, making it one of the most heavily funded areas of medical research today.

Angiogenesis is an essential process during development — growth of a vascular system is one of the earliest events in organogenesis. Nonetheless it also occurs in adulthood, during wound healing and restoration of blood flow to injured tissues.

Angiogenesis is regulated by a very sensitive interplay of growth factors and inhibitors, and their imbalance can lead to disease. In cancer, diabetic eye disease and rheumatoid arthritis, excessive angiogenesis feeds diseased tissue and destroys normal tissue. Conversely, insufficient angiogenesis underlies conditions such as coronary heart disease, stroke and delayed wound healing, where inadequate blood-vessel growth leads to poor circulation and tissue death.

This Insight describes many of these physiological and pathophysiological processes of angiogenesis and lymphangiogenesis (the development of new lymph vessels) from development through to the immune response and nervous system function. In addition, it introduces some exciting therapeutic applications that have recently been made available. We are indebted to all our authors.

We are pleased to acknowledge the support from our corporate sponsor Genentech and academic sponsors the NIH and the JDRF in producing this Insight. As always, *Nature* retains sole responsibility for editorial content and peer review.

Natalie DeWitt, Senior Editor

REVIEWS

932 Angiogenesis in life, disease and medicine

P. Carmeliet

937 Endothelial cells and VEGF in vascular development

L. Coultas,
K. Chawengsaksophak &
J. Rossant

946 Lymphangiogenesis in development and human disease

K. Alitalo, T. Tammela &
T. V. Petrova

954 From angiogenesis to neuropathology

D. A. Greenberg &
K. Jin

960 Retinal angiogenesis in development and disease

R. F. Gariano &
T. W. Gardner

967 Angiogenesis as a therapeutic target

N. Ferrara &
R. S. Kerbel

nature
insight

Angiogenesis in life, disease and medicine

Peter Carmeliet¹

The growth of blood vessels (a process known as angiogenesis) is essential for organ growth and repair. An imbalance in this process contributes to numerous malignant, inflammatory, ischaemic, infectious and immune disorders. Recently, the first anti-angiogenic agents have been approved for the treatment of cancer and blindness. Angiogenesis research will probably change the face of medicine in the next decades, with more than 500 million people worldwide predicted to benefit from pro- or anti-angiogenesis treatments.

Blood vessels arose during evolution to carry oxygen to distant organs. Not surprisingly, these vessels are crucial for organ growth in the embryo and repair of wounded tissue in the adult. But an imbalance in the growth of blood vessels contributes to the pathogenesis of numerous disorders. In less than 15 years, an explosion of interest in angiogenesis research has generated the necessary insights to develop the first clinically approved anti-angiogenic agents.

Here I discuss some key mechanisms of angiogenesis and opportunities to develop further novel therapeutic strategies that target this process, to minimize the adverse effects of these treatments and to avoid resistance to this novel medicine.

The discovery of blood and lymph vessels

In primitive animals, such as the worm *Caenorhabditis elegans* and the fruitfly *Drosophila melanogaster*, oxygen is capable of diffusing throughout their small body to all cells. In other species, which developed later in evolution and grew to larger sizes, a vascular network distributes oxygen in the blood to distant cells. The Ancient Greek physician Galen originally proposed that the blood does not circulate but is locally regenerated by the body when its supplies are consumed. Only in 1628 did William Harvey discover that the heart pumps the blood around the body through arteries and that veins return the blood to the heart. A few decades later in 1661, Marcello Malpighi identified the capillaries as the smallest vessels that close the circulatory loop between arteries and veins (Fig. 1a). Around the same time, Caspar Aselius discovered another type of vessel, the lymphatic vessel. Because of the blood pressure, blood plasma continuously leaks from the capillaries, and lymph vessels return this fluid back to the blood circulation. Although blood vessels arose earlier in evolution, lymph vessels are only present in amphibians onwards¹ (Fig. 1b).

The first vessels in life

In the embryo, blood vessels provide the growing organs with the necessary oxygen to develop. Apart from their nutritive function, vessels also provide instructive trophic signals to promote organ morphogenesis (see the review by Coultas, Chawengsaksophak and Rossant in this issue, p. 957). Blood vessels arise from endothelial precursors, which share an origin with haematopoietic progenitors. This close link between the blood and blood vascular systems remains important for angiogenesis throughout life, even in disease (see below). These progenitors assemble into a primitive vascular labyrinth of small capillaries — a process known as vasculogenesis (Fig. 1c). Interestingly, already at this stage capillaries have acquired an arterial and venous cell fate, indi-

cating that vascular-cell specification is genetically programmed and not only determined by haemodynamic force (see p. 937).

During the angiogenesis phase, the vascular plexus progressively expands by means of vessel sprouting and remodels into a highly organized and stereotyped vascular network of larger vessels ramifying into smaller ones (Fig. 1c). Nascent endothelial-cell (EC) channels become covered by pericytes (PCs) and smooth muscle cells (SMCs), which provide strength and allow regulation of vessel perfusion, a process termed arteriogenesis. As reviewed by Alitalo, Tammela and Petrova in this issue (p. 946), the lymphatic system develops differently, as most lymphatics transdifferentiate from veins.

Over the past 15 years, genetic studies in mice, zebrafish and tadpoles have provided insights into the key mechanisms and molecular players that regulate the growth of blood vessels (angiogenesis) or lymph vessels (lymphangiogenesis) in the embryo (see p. 937 and p. 946). For instance, members of the Notch family drive the arterial gene programme, whereas the orphan receptor COUP-TFII regulates venous specification. The homeobox gene *Prox-1*, by contrast, is a master switch of lymphatic commitment. VEGF and its homologue VEGF-C are key regulators of vascular and lymphatic EC sprouting, respectively, whereas platelet-derived growth factor (PDGF)-BB and angiopoietin-1 recruit mural cells around endothelial channels. The formation of vessels is a complex process, requiring a finely tuned balance between numerous stimulatory and inhibitory signals, such as integrins, angiopoietins, chemokines, junctional molecules, oxygen sensors, endogenous inhibitors and many others². An exciting recent development is the discovery of the links between vessels and nerves and, in particular, how axon-guidance signals such as Ephrins, Semaphorins, Netrins and Slits allow vessels to navigate to their targets or control vessel morphogenesis³. Angiogenic signals also guide axons and affect neurons in health and disease, as reviewed by Greenberg in this issue (p. 954).

Vessels of disease and death

After birth, angiogenesis still contributes to organ growth but, during adulthood, most blood vessels remain quiescent and angiogenesis occurs only in the cycling ovary and in the placenta during pregnancy. However, ECs retain their remarkable ability of dividing rapidly in response to a physiological stimulus, such as hypoxia for blood vessels and inflammation for lymph vessels² (see p. 946). As such, (lymph)angiogenesis is reactivated during wound healing and repair. But in many disorders, this stimulus becomes excessive, and the balance between stimulators and inhibitors is tilted, resulting in a (lymph)angiogenic switch. The best-known conditions in which angiogenesis is

¹Center of Transgene Technology and Gene Therapy, University of Leuven, Flanders Interuniversity Institute for Biotechnology (VIB), B-3000 Leuven, Belgium.

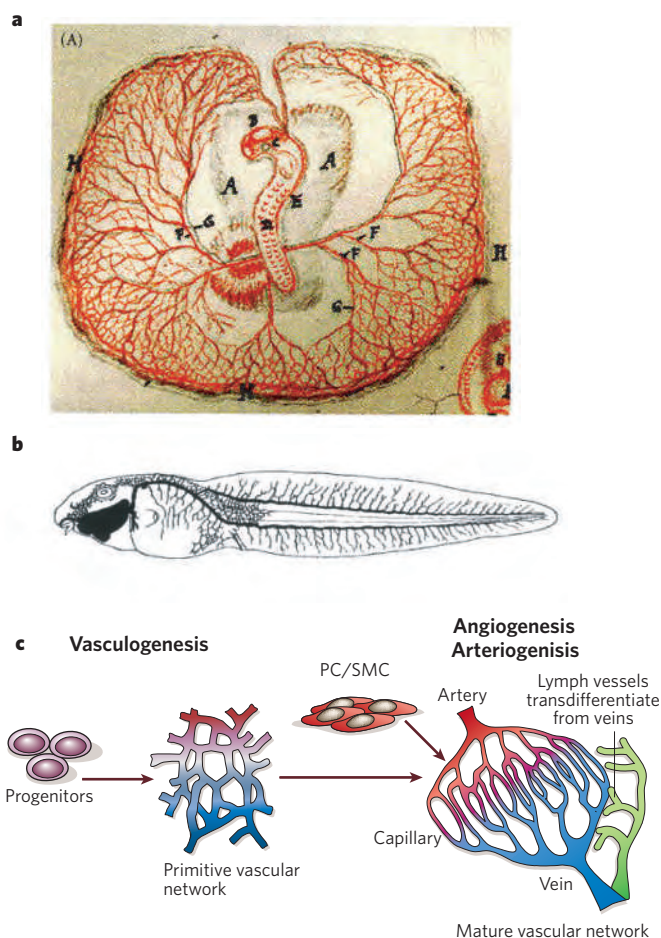


Figure 1 | History and formation of blood and lymph vessels. **a**, Drawing by M. Malpighi (1661), showing the vascular network of arteries, capillaries and veins in a developing chicken embryo (from ref. 26). **b**, Drawing by M. Hoyer displaying the lymphatic network in early tadpoles (from ref. 27). **c**, Development of the vascular systems: during vasculogenesis, endothelial progenitors give rise to a primitive vascular labyrinth of arteries and veins; during subsequent angiogenesis, the network expands, pericytes (PCs) and smooth muscle cells (SMCs) cover nascent endothelial channels, and a stereotypically organized vascular network emerges. Lymph vessels develop via transdifferentiation from veins.

switched on are malignant, ocular and inflammatory disorders, but many additional processes are affected, such as obesity, asthma, diabetes, cirrhosis, multiple sclerosis, endometriosis, AIDS, bacterial infections and autoimmune disease (for a more complete list see Supplementary Table 1). There is even a close link between angiogenesis, neural stem cells and learning.

In other diseases, such as ischaemic heart disease or preeclampsia, the angiogenic switch is insufficient, causing EC dysfunction, vessel malformation or regression, or preventing revascularization, healing and regeneration (for a more complete list see Supplementary Table 2). Besides its vascular activity, VEGF is also trophic for nerve cells, lung epithelial cells and cardiac muscle fibres, further explaining why insufficient VEGF levels contribute to neurodegeneration⁴, respiratory distress and, possibly, cardiac failure (Supplementary Table 2). Angiogenesis has been implicated in more than 70 disorders so far, and the list is ever growing. In this issue, Gariano and Gardner, and Ferrara and Kerbel discuss the key signals of (lymph)angiogenesis in pathological conditions (see pp. 960 and 967. See also p. 946). Interestingly, some molecules such as PlGF (a homologue of VEGF) have a role in angiogenesis in disease without affecting quiescent vessels in healthy organs, making them attractive therapeutic targets for the development of safe anti-angiogenic drugs⁵.

Angiogenesis as a promising medicine

Over the past decade, intensive efforts have been undertaken to develop therapeutic strategies to promote revascularization of ischaemic tissues or to inhibit angiogenesis in cancer, ocular, joint or skin disorders. Unfortunately, clinical trials testing the pro-angiogenic potential of VEGF or fibroblast growth factor (FGF) have not had the expected results⁶. Although part of this failure is attributable to suboptimal delivery strategies, stimulating the growth of durable and functional vessels is a more formidable challenge than previously anticipated. Novel strategies, involving transplantation of bone-marrow-derived cells or the delivery of molecules capable of stimulating the growth not only of distal capillaries but also of proximal collateral conduit vessels, may be required in the future⁶. Stimulating lymphangiogenesis is also emerging as a novel treatment of lymphoedema (see p. 946).

Angiogenesis does not initiate malignancy but promotes tumour progression and metastasis. Unlike tumour cells, ECs are genomically stable and were therefore originally considered to be ideal therapeutic targets that would not become resistant to anti-angiogenic therapy. Most previous efforts have thus been focused on developing anti-angiogenic agents that primarily target ECs. Several reviews in this issue and a forthcoming review by Jain and colleagues⁷ provide an update on the clinical use of the first two FDA-approved VEGF antagonists in ocular and malignant disease and discuss opportunities to inhibit lymphangiogenesis in cancer. The recent clinical experience with VEGF inhibitors has provided a number of important, but puzzling, insights and raised various outstanding questions. First, the anti-VEGF antibody Avastin (Genentech) only provides an overall survival benefit in colorectal, breast and lung cancer patients when combined with conventional chemotherapy. It is still not entirely clear why anti-VEGF monotherapy was ineffective in humans, whereas it was effective in rodents. Second, monotherapy with the multi-targeted receptor tyrosine kinase inhibitors (RTKIs) Sorafenib (Bayer and Onyx) or Sutent (Pfizer), which target ECs as well as cancer, and probably also stromal and haematopoietic cells, demonstrates clinical benefit in certain cancers. But does this imply that future anti-angiogenic strategies should target both ECs and non-EC types? Third, despite its ability to block all three VEGF receptors, Vatalanib (Novartis and Schering AG) does not substantially enhance the benefit of conventional chemotherapy. How can this discrepancy with Avastin be explained? Fourth, despite promising success, cancer patients receiving a single class of angiogenesis inhibitors, even in combination with chemotherapy, still die. Does this suggest that the anti-angiogenic strategy is insufficient or does it evoke resistance and, if so, how can we avoid resistance? Can we develop more reliable biomarkers to monitor the efficacy of anti-angiogenic therapy⁷ (p. 967)? Fifth, adverse effects have been reported. What are the molecular mechanisms of these effects? For reasons of brevity, I highlight here only a few key issues. Taking the stand that targeting ECs or the principal angiogenic factor VEGF alone may not (ever) suffice to eradicate malignant tumours, I discuss here alternative options to complement the current VEGF-based therapies with strategies that target other angiogenic factors or target, in combination, other non-EC types that indirectly affect angiogenesis (Box 1). In addition, possible mechanisms of the adverse effects and resistance to anti-angiogenic therapy will be highlighted (Box 2).

Inhibition of angiogenesis by targeting endothelial cells

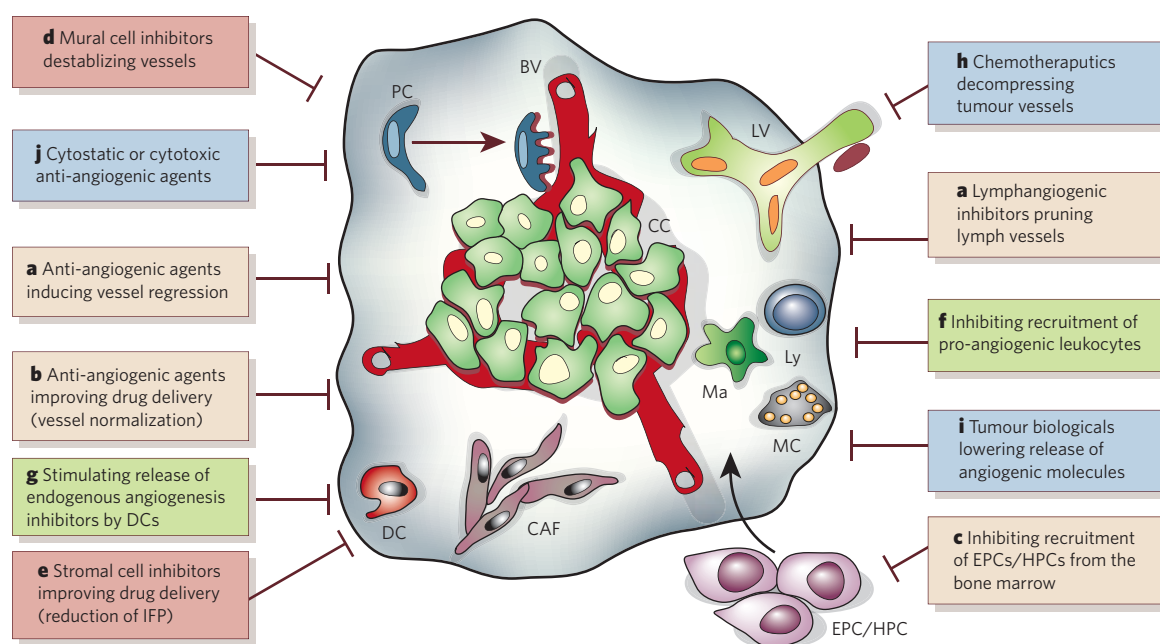
The best-known anti-angiogenic agents of this class are the VEGF inhibitors. The most advanced in the clinic are the anti-VEGF antibody Avastin, a VEGF165 aptamer (Macugen, Eyetech) and various RTKIs, which target VEGFRs and other receptors (see p. 967 and ref. 7). Additional compounds targeting VEGF family members, currently in development, include a VEGF trap (Regeneron) and antibodies against VEGFR-2 or VEGFR-1 (Imclone) and against the VEGFR-1 ligand PlGF (Thrombogenics Ltd and BioInvent International). This class of anti-angiogenic agents not only arrests EC proliferation and prevents vessel growth, but also induces regression of existing vessels by increasing EC death (Box 1). Immature vessels,

Box 1 | Strategies targeting endothelial and non-endothelial cells to inhibit tumour angiogenesis

Tumour angiogenesis has classically been inhibited by anti-angiogenic agents that affect ECs directly. Alternative anti-tumour angiogenesis strategies target other cell types in tumours (mural and stromal cells, haematopoietic cells and tumour cells), which stimulate angiogenesis indirectly. The yellow boxes show agents (such as VEGF inhibitors, metronomic chemotherapy and other compounds) that target endothelial (progenitor) cells (EPCs); they inhibit (lymph)angiogenesis (**a**), induce vessel regression (**a**) and normalization (**b**), and block

recruitment of EPCs (**c**). The red boxes show agents (such as PDGF inhibitors) that target mural and stromal cells and destabilize vessels (**d**), reduce the release of pro-angiogenic factors or progenitor cytokines, and lower the interstitial fluid pressure (IFP), which improves drug delivery (**e**). The green boxes indicate agents (such as VEGFR-1 inhibitors, chemokine antagonists and so on) that target haematopoietic cells and reduce the infiltration of pro-angiogenic bone-marrow-derived precursors and mature leukocytes (**c,f**), and stimulate the release of endogenous

angiogenesis inhibitors in dendritic cells (**g**). The blue boxes show agents targeting cancer cells (chemotherapy, radiation, tumour-cell-targeted biologicals) that improve drug delivery by decompressing tumour vessels (**h**) and decrease the release of (lymph)angiogenic factors (**i**); some anti-angiogenic agents are also cytotoxic for tumour cells (**j**). BV, blood vessel; CAF, carcinoma-activated fibroblast; CC, cancer cell; DC, dendritic cell; LV, lymph vessel; Ly, lymphocyte; Ma, macrophage; PC, pericyte; MC, mast cells.



devoid of pericytes, are most susceptible. In addition, VEGF inhibitors suppress the mobilization of endothelial progenitor cells (EPCs) from the bone marrow. Anti-VEGF treatment also improves cytotoxic drug delivery by normalizing the chaotic pattern and abnormal architecture of tumour vessels, and reducing vascular permeability and the interstitial fluid pressure, explaining why this antibody acts as a chemosensitizer and increases the efficacy of chemotherapeutics⁸. Besides these anti-EC activities, VEGF inhibitors are cytotoxic for some malignant cells, activate the anti-tumour immune attack and suppress the pro-angiogenic activity of haematopoietic cells (see below; Box 1). Chemotherapeutics, which target dividing cells, may also inhibit EC growth when delivered in metronomic regimens (that is, a continuous low dose)^{7,9} (see p. 967).

Additional anti-angiogenic agents are currently being evaluated. However, the potential to combine VEGF antagonists with other inhibitors of distinct angiogenic targets remains largely untested in the clinic, despite emerging evidence that many more angiogenic factors besides VEGF contribute to the angiogenic switch in tumours, especially in the advanced stage. Nonetheless, such anti-angiogenic combination therapy might very well increase the efficacy of and decrease the resistance to angiogenesis inhibition.

Angiogenesis inhibition by targeting mural and stromal cells

Vessels in tumours are covered by PCs¹⁰. These mural cells differentiate from pools of c-Kit⁺Sca-1⁺VEGFR-1⁺ perivascular progenitor cells, which are mobilized from the bone marrow in response to PDGF-BB¹¹. By locally releasing VEGF (an EC survival factor) and angiopoietin-1

(which tightens vessels by means of a matrix and cell–cell contacts), these mural cells promote vessel stabilization. When PDGFs are overexpressed, tumour vessels are covered by more mural cells and tumour growth is accelerated¹⁰. Conversely, when PDGFR β signaling is inhibited, fewer PCs are recruited, tumour vessels are dilated and EC apoptosis is increased. Combined administration of RTKIs, targeting VEGFRs and PDGFR β , increases the anti-angiogenic effect, even in the often-intractable late stage of solid tumours¹². PDGFR β inhibitors also destabilize the larger SMC-covered vessels, which supply bulk flow to tumours and render them more susceptible to EC-specific inhibitors (Box 1).

In 1889, Paget proposed that 'seeds' of tumour cells form metastatic deposits only if they land in appropriate 'soils'. The reactive tumour stroma is not simply an innocent bystander but an active contributor to tumour progression. Unlike normal tissue, the tumour stroma contains inflammatory infiltrates, an increased microvessel density and dysfunctional lymphatics, a different and more dense extracellular matrix (ECM) and carcinoma-activated fibroblasts (CAFs). CAFs accelerate tumour growth and may increase their malignancy; they also affect the tumour vasculature in many ways. Indeed, CAFs express PDGFR β and are recruited to the tumour, proliferate and release angiogenic factors such as VEGF and PlGF in response to PDGF-BB¹⁰. As well as suppressing angiogenesis, PDGF-BB antagonists lower the interstitial fluid pressure and improve drug delivery through the tumour vasculature (Box 1). Although the precise reason *in vivo* remains speculative, CAFs exert a tension on microfibrillar networks *in vitro* and, when stimulated by PDGF-BB, contract the interstitial matrix, thereby compressing tumour vessels¹³. Stromal

fibroblasts also recruit EPCs by releasing stromal-cell-derived factor 1 α (SDF-1 α)¹⁴. Inhibiting this chemokine also inhibits tumour growth (Box 1). Unanswered questions are whether the stroma also renders ECs resistant to chemotherapeutics through cell-adhesion-mediated drug resistance (CAM-DR)¹⁵ and whether stromal cells provide niches for cancer (or endothelial) stem cells.

Inhibition of angiogenesis by targeting haematopoietic cells

In 1863, Virchow postulated that inflammation stimulates the progression of cancer. Tumour cells produce various cytokines and chemokines that attract macrophages, dendritic cells, mast cells, T cells and haematopoietic progenitors. Tumour-derived VEGF and PlGF also recruit and stimulate the survival of some of these cells. Apart from releasing mitogenic and survival factors for tumour and stromal cells, stimulating DNA damage, facilitating invasion by means of ECM remodelling and evading the host defence, inflammatory cells also stimulate (lymph)angiogenesis in tumours¹⁶ (Box 1). For instance, tumour-associated macrophages (TAMs) accumulate in hypoxic tumour regions and produce (lymph)angiogenic factors such as VEGF, VEGF-C and VEGF-D. Tie2-expressing monocytes (TEMs),

mast cells and platelets also release pro-angiogenic factors¹⁷. Certain leukocyte-attracting chemokines such as IL-8 directly stimulate EC growth; inhibiting this chemokine retards tumour growth. Blocking the signals that promote leukocyte infiltration and survival may thus inhibit tumour angiogenesis.

In the embryo, haematopoietic stem cells (HSCs) migrate into avascular areas and attract sprouting ECs by releasing angiogenic factors, such as angiopoietin-1 (ref. 18). In the adult, bone-marrow-derived haematopoietic cells expressing markers such as Sca-1, c-Kit, CXCR4 and/or VEGFR-1 become recruited, often together with EPCs, to tumours or ischaemic tissues in response to VEGF and PlGF^{14,19,20}. These angio-competent cells extravasate around nascent vessels, where they are retained by SDF-1 α , and stimulate growth of resident vessels by releasing angiogenic factors such as VEGF, PlGF and angiopoietin-2 (Box 1). In other cases, these cells function as haemangioblasts, producing both haematopoietic and endothelial progenitors that give rise to new blood vessels. Moreover, in response to PlGF released by tumour cells, VEGFR-1⁺ haematopoietic bone-marrow progenitors home to tumour-specific premetastatic sites, where they recruit tumour cells and EPCs; anti-VEGFR-1 antibodies prevent the forma-

Box 2 | Mechanism of acquired resistance to anti-angiogenic agents

Despite the promising successes of anti-VEGF therapy, cancer patients still die. Emerging evidence suggests that this may be due, at least in part, to acquired resistance to anti-angiogenic agents. Several possible mechanisms are highlighted.

Tumour-cell-related mechanisms are shown below in blue. **a**, During tumour progression, mutant tumour cell clones (yellow and blue cancer cells) may become selected and express more of the same or other angiogenic factors. Tumour cells may also upregulate additional angiogenic factors in response to anti-angiogenic treatment (that is upregulation of PlGF and FGF-2 after VEGF inhibition, for example, of VEGF after EGFR or VEGFR-2 inhibition or IL-8 after HIF-1 inhibition). **b**, Mutant tumour cell clones (for instance, those lacking p53 or HIF-1) or pro-angiogenic

inflammatory cells may survive better in hypoxic tumours after angiogenesis inhibition; their reduced vascular dependence impairs the anti-angiogenic response. **c**, Some RTKs do not synergize with chemotherapy.

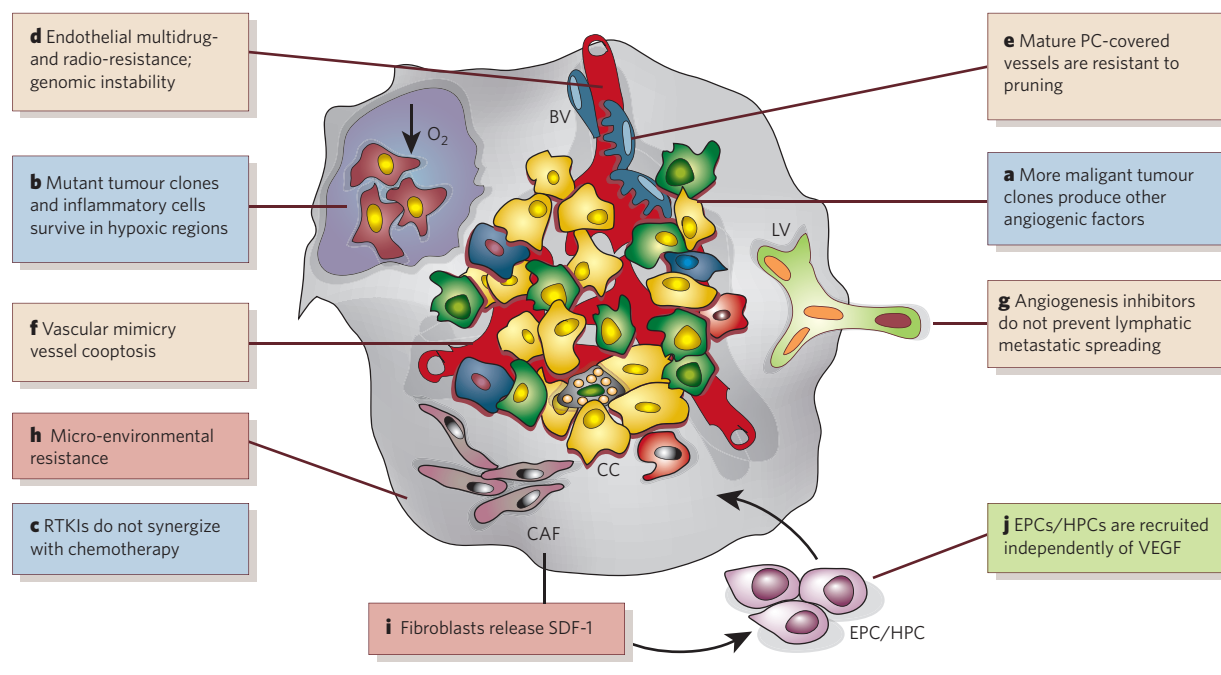
EC-related mechanisms are shown below in yellow. **d**, ECs are chemoprotected by high levels of VEGF and other EC survival factors in tumours, which upregulate anti-apoptotic signals and multidrug-resistance-associated proteins. Hypoxic activation of HIF-1 also renders ECs resistant to irradiation. In rare cases, ECs even exhibit cytogenetic abnormalities and may be genomically unstable, but this has only been observed in some human cancers. **e**, Pre-existing supply vessels are covered by SMCs and are not easily pruned by EC-targeted treatment. **f**, A fraction of tumour vessels, lined by malignant cells (vascular mimicry) or co-opted from existing vessels, may be less sensitive to anti-

angiogenic treatment. **g**, Tumour cells metastasize through lymph vessels; their growth is not (necessarily) blocked by anti-angiogenic therapy.

Stromal-cell-related mechanisms are shown below in red. **h**, Tissue-specific differences in the micro-environment may determine tumour malignancy (for instance, HIF-1^{-/-} gliomas are more malignant in the brain than the skin).

i, CAFs produce SDF-1 to recruit EPCs/HPCs.

Bone-marrow-derived cell-related mechanisms are shown below in green. **j**, Tumours recruit pro-angiogenic EPCs, HPCs and inflammatory cells independently of VEGF. PC, pericyte; BV, blood vessel; LV, lymph vessel; CC, cancer cell; CAF, carcinoma-activated fibroblast; SDF-1, stromal-cell-derived factor 1. References and additional (more hypothetical) mechanisms are listed in Supplementary Table 3.



tion of such premetastatic niches²¹. Blocking mobilization of these cells by interfering with SDF-1 and PlGF is thus a novel strategy to reduce tumour angiogenesis, growth and even metastasis. Broad-spectrum RTKIs, which block c-Kit, may also have similar effects. Stimulating the function of dendritic cells (DCs) might also be considered, as these antigen-presenting cells not only mount an anti-tumour immune attack but also release endogenous anti-angiogenic cytokines. Because PlGF and VEGFR-1 suppress DC function²², inhibiting these molecules offers novel opportunities (Box 1).

Inhibition of angiogenesis by targeting neoplastic cells

Future standard treatment of cancers will involve the use of cytotoxic, radiation or tumour-cell-targeted biological tools to destroy malignant cells. These regimens also inhibit tumour angiogenesis, directly or indirectly (Box 1). Indeed, blood and lymph vessels often collapse under the high compressive mechanical stresses inside tumours. By destroying tumour cells, vessels are decompressed, resulting in increased perfusion and drug delivery. Furthermore, tumour cells release numerous angiogenic molecules and induce the expression of angiogenic receptors in tumour vessels (for instance, EGF induces endothelial growth factor (EGF) receptors and VEGFRs in tumour-associated vessels) (see p. 967). Thus, Herceptin (Genentech), an anti-EGFR antibody used to block the growth of neoplastic epithelial cells, also acts as an anti-angiogenic cocktail by lowering angiogenic factors and upregulating endogenous angiogenesis inhibitors²³. Many other EGFR inhibitors could have similar anti-angiogenic activities⁷. By producing factors that induce lymph node lymphangiogenesis, primary tumours also prepare their future lymphatic metastatic transport to sentinel lymph nodes²⁴.

Anti-angiogenic factors may also be cytostatic or cytotoxic for tumour cells (Box 1). Indeed, tumour cells often express receptors for VEGF (VEGFR-1 and Neuropilin1), PDGF, FGF, EGF, stem-cell factor (SCF) and other angiogenic factors²⁵ (see p. 967). Hence, anti-angiogenic drugs could lead to the direct killing of cancer cells by interfering with survival pathways and/or enhancing sensitivity to other treatments. The potential of the broad-spectrum RTKIs Sorafenib or Sutent to inhibit both EC and tumour-cell division (and possibly also that of other non-ECs) may explain their efficacy as monotherapy for renal cell carcinoma and gastrointestinal stromal tumours, respectively. Of interest, Neuropilin1 mediates VEGF-driven survival and migration of tumour cells but lacks a tyrosine kinase domain, and is thus not inhibited by RTKIs.

Conclusions and future directions

Angiogenesis inhibitors are likely to change the face of medicine in the next decade. Because of VEGF's predominant role in angiogenesis, inhibition of VEGF seems to be necessary but is probably insufficient to permanently halt this process in many disorders. In fact, emerging evidence indicates that inhibition of a single target leads to upregulation of additional angiogenic factors: for instance, PlGF is upregulated after anti-VEGF therapy, VEGF after anti-VEGFR-2 or anti-EGFR administration, and interleukin (IL)-8 after hypoxia-inducible factor 1 (HIF-1) deletion (see Supplementary Table 3 for more information). Combined treatment of anti-angiogenic agents with distinct complementary mechanisms of action, targeting other angiogenic molecules and/or targeting not only ECs but also other pro-angiogenic cell types, may thus offer advantages of increased efficacy — at least if toxicity is not a concern (see below). Another advantage is that such combinations may give the tumour less chance to escape from anti-angiogenic treatment. Exploring strategies to delay, minimize or even avoid resistance to anti-angiogenic agents might further increase the benefit of anti-angiogenic treatments. A number of known and hypothetical mechanisms of resistance to anti-angiogenesis are listed in Box 2 and Supplementary Table 3.

As anti-angiogenic agents are likely to be delivered earlier and earlier to more and more patients for less advanced, life-threatening disease, probably in combination with additional medications, the safety of anti-angiogenic treatment is a topic of emerging importance. On the

basis of pharmacological and genetic studies in mice, inhibition of VEGF-driven angiogenesis might have been expected to cause many more adverse effects (Supplementary Table 4). Fortunately, such toxicity has not been observed in humans, but it may emerge in conditions where the risk is increased by genetic predisposition or pharmacological treatment. Some of the adverse effects of anti-VEGF therapy can be explained by the requirement of threshold levels of VEGF for the survival and maintenance of quiescent vessels in healthy organs (Supplementary Table 4). An attractive, novel class of target thus includes molecules, such as PlGF, that only affect angiogenesis in disease without affecting quiescent vessels in healthy organs⁵. The challenge for the future is to develop such novel anti-angiogenic strategies and to optimize combinatorial treatment regimens to fully exploit the therapeutic potential of angiogenesis inhibition. ■

1. Ny, A. et al. A genetic *Xenopus laevis* tadpole model to study lymphangiogenesis. *Nature Med.* **11**, 998–1004 (2005).
2. Carmeliet, P. Angiogenesis in health and disease. *Nature Med.* **9**, 653–660 (2003).
3. Carmeliet, P. & Tessier-Lavigne, M. Common mechanisms of nerve and blood vessel wiring. *Nature* **436**, 193–200 (2005).
4. Lambrechts, D., Störkebaum, E. & Carmeliet, P. VEGF: necessary to prevent motoneuron degeneration, sufficient to treat ALS? *Trends Mol. Med.* **10**, 275–282 (2004).
5. Lutun, A., Autiero, M., Tjwa, M. & Carmeliet, P. Genetic dissection of tumour angiogenesis: are PlGF and VEGFR-1 novel anti-cancer targets? *Biochim. Biophys. Acta* **1654**, 79–94 (2004).
6. Simons, M. Angiogenesis: where do we stand now? *Circulation* **111**, 1556–1566 (2005).
7. Jain, R. K., Duda, D. G., Clark, J. W. & Loeffler, J. S. Lessons from phase III clinical trials on anti-VEGF therapy for cancer. *Nature Clin. Pract. Oncol.* (in the press).
8. Jain, R. K. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* **307**, 58–62 (2005).
9. Kerbel, R. S. & Kamen, B. A. The anti-angiogenic basis of metronomic chemotherapy. *Nature Rev. Cancer* **4**, 423–436 (2004).
10. Ostman, A. PDGF receptors—mediators of autocrine tumor growth and regulators of tumor vasculature and stroma. *Cytokine Growth Factor Rev.* **15**, 275–286 (2004).
11. Song, S., Ewald, A. J., Stallcup, W., Werb, Z. & Bergers, G. PDGFR β + perivascular progenitor cells in tumours regulate pericyte differentiation and vascular survival. *Nature Cell Biol.* **7**, 870–879 (2005).
12. Bergers, G., Song, S., Meyer-Morse, N., Bergsland, E. & Hanahan, D. Benefits of targeting both pericytes and endothelial cells in the tumor vasculature with kinase inhibitors. *J. Clin. Invest.* **111**, 1287–1295 (2003).
13. Heldin, C. H., Rubin, K., Pietras, K. & Ostman, A. High interstitial fluid pressure — an obstacle in cancer therapy. *Nature Rev. Cancer* **4**, 806–813 (2004).
14. Orimo, A. et al. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**, 335–348 (2005).
15. Buttery, R. C., Rintoul, R. C. & Sethi, T. Small cell lung cancer: the importance of the extracellular matrix. *Int. J. Biochem. Cell Biol.* **36**, 1154–1160 (2004).
16. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
17. De Palma, M. et al. Tie2 identifies a hematopoietic lineage of proangiogenic monocytes required for tumor vessel formation and a mesenchymal population of pericyte progenitors. *Cancer Cell* **8**, 211–226 (2005).
18. Takakura, N. et al. A role for hematopoietic stem cells in promoting angiogenesis. *Cell* **102**, 199–209 (2000).
19. Rafii, S. & Lyden, D. Therapeutic stem and progenitor cell transplantation for organ vascularization and regeneration. *Nature Med.* **9**, 702–712 (2003).
20. Grunewald, M. et al. VEGF-induced adult neovascularization depends on SDF-1-mediated retention of bone marrow derived accessory cells. *Cell* (in the press).
21. Kaplan, R. N. et al. VEGFR1-positive hematopoietic bone marrow progenitors initiate the premetastatic niche. *Nature* **438**, 820–827 (2005).
22. Zou, W. Immunosuppressive networks in the tumour environment and their therapeutic relevance. *Nature Rev. Cancer* **5**, 263–274 (2005).
23. Izumi, Y., Xu, L., di Tomaso, E., Fukumura, D. & Jain, R. K. Tumour biology: Herceptin acts as an anti-angiogenic cocktail. *Nature* **416**, 279–280 (2002).
24. Hirakawa, S. et al. VEGF-A induces tumor and sentinel lymph node lymphangiogenesis and promotes lymphatic metastasis. *J. Exp. Med.* **201**, 1089–1099 (2005).
25. Fletcher, J. A. Role of KIT and platelet-derived growth factor receptors as oncoproteins. *Semin. Oncol.* **31**, 4–11 (2004).
26. Gilbert, S. F. *Developmental Biology*, 6th edn (Swarthmore College, Sinauer Assoc., Sunderland, MA, 2000).
27. Hoyer, M. Untersuchungen ueber das Lymphgefäßsystem der Froschlärven. *Bull. Acad. Cracov. Teil II*, 451–464 (1905).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The author regrets that, owing to space limitations, he has been unable to refer to all of the primary literature and had to rely instead, in many instances, on reviews. P.C. is supported by grants from FWO, the European Union and the Concerted Research Activities of Belgium.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany this paper at www.nature.com/nature. Correspondence should be addressed to P.C. (peter.carmeliet@med.kuleuven.be).

Endothelial cells and VEGF in vascular development

Leigh Coultas¹, Kallayane Chawengsaksophak¹ & Janet Rossant¹

The intricate patterning processes that establish the complex vascular system during development depend on a combination of intrinsic pre-patterning and extrinsic responses to environmental parameters. Mutational studies in mice and fish have shown that the vascular system is highly sensitive to genetic disruption and have identified potential targets for therapeutic interventions. New insights into non-vascular roles of vascular endothelial growth factor and the requirement for endothelial cells in adult organs and stem-cell niches highlight possible side effects of anti-angiogenic therapy and the need for new targets.

The development of the vascular system is one of the earliest events in organogenesis. The early blood vessels of the embryo and yolk sac in mammals develop by aggregation of *de-novo*-forming angioblasts into a primitive vascular plexus (vasculogenesis), which then undergoes a complex remodelling process, in which growth, migration, sprouting and pruning lead to the development of a functional circulatory system (angiogenesis; Fig. 1). Many of the events that occur during the normal progression of vascular development in the embryo are recapitulated during situations of neoangiogenesis in the adult¹. Most notably, many tumours promote their own growth and dispersion to form metastases by recruiting host blood vessels to grow into the vicinity of the tumour (so-called tumour angiogenesis). In addition, neovascularization induced after tissue damage is a key component of the repair and healing process.

In recent years, there have been major breakthroughs in our understanding of the genetic control of the normal processes of vascular development and remodelling, especially from the characterization of vascular-mutant phenotypes in mice². Embryonic phenotypes that fail to develop different phases of the normal vasculature have been reported for the vascular endothelial growth factor (VEGF)/VEGF receptor (VEGFR) and angiopoietin/Tie families of vascular-specific signalling molecules. Similar results were reported for certain members of more widely used signalling pathways including platelet-derived growth factor (PDGF), transforming growth factor- β (TGF- β), ephrin, Notch and pathways thought to be primarily involved in axon guidance (such as semaphorins, netrins and Robo). Excitingly, many of these signalling pathways are reactivated in situations of neoangiogenesis³. This finding opens up a new era of rational therapeutics for prevention of tumour angiogenesis or promotion of new blood-vessel growth.

In this review, we focus on recent studies in fish, chicks and mice that have dissected early patterning of the vascular endothelial system, with particular reference to the role of endothelial cells and VEGF. Later events and signalling pathways that are equally important in vascular development, such as PDGF, TGF- β and angiopoietin signalling in pericyte and smooth-muscle-cell recruitment and vascular remodelling, have been reviewed elsewhere^{4–6}. The process of lymphangiogenesis is reviewed in this issue by Alitalo, Tammela and Petrova (p. 946). Furthermore, a comprehensive list of genes that

are essential for vascular development has been compiled and is given in Supplementary Table 1.

Endothelial cells might probably have more interesting roles than just acting as channels for blood circulation. They can promote stem-cell development and organ formation by acting as sources of

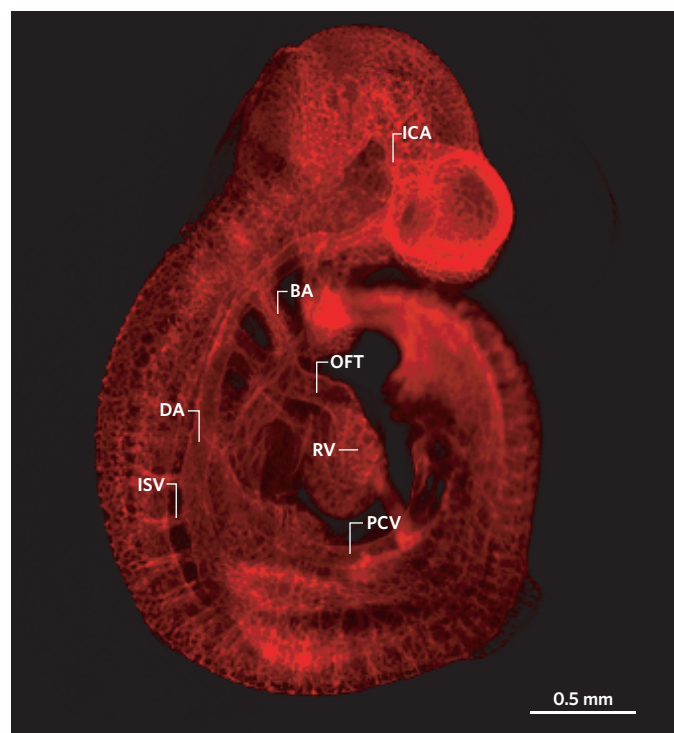


Figure 1 | Murine embryonic vasculature. Projection from optical projection tomography¹⁰⁰ of a normal mouse embryo at embryonic day (E) 9.5 showing developing vasculature labelled with CD31 (PECAM) immunofluorescence staining (K.C. and J.R., unpublished data). BA, branchial arteries; DA, dorsal aorta; ICA, intercarotid artery; ISV, intersomitic vessels; OFT, outflow tract; PCV, posterior cardinal vein; RV, right ventricle. (Image courtesy of L. Davidson, Mouse Imaging Centre, Hospital for Sick Children, Toronto, Canada.)

¹The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, M5G 1X8 Canada.

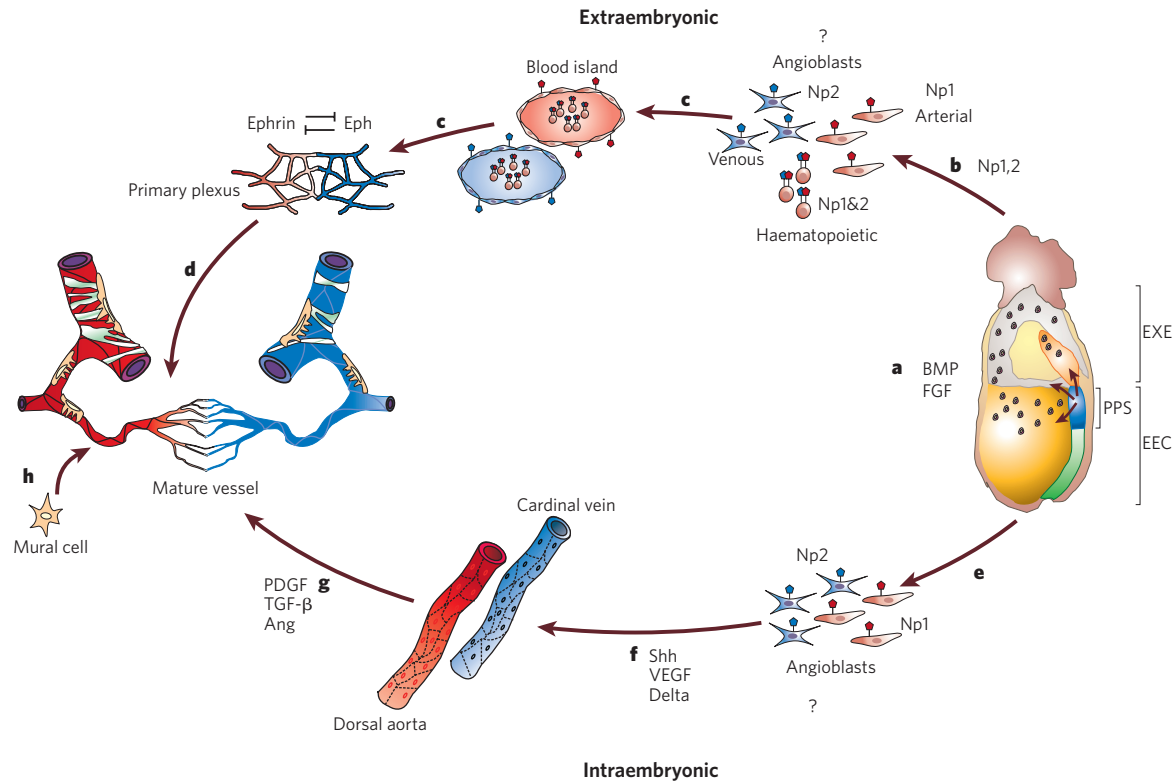


Figure 2 | Formation of a functional circulation from endothelial progenitors. **a**, Vascular progenitors appear in response to basic fibroblast growth factor (bFGF) and bone morphogenetic protein 4 (BMP4) in the posterior primitive streak (PPS) as vascular endothelial growth receptor-2 (VEGFR-2)/Flk1-positive mesodermal cells. **b**, Flk1-positive cells in the primitive streak give rise to both blood and endothelium (haemangioblasts), but are restricted to haematopoietic or angiogenic fate after emigrating into extra-embryonic sites (extra-embryonic ectoderm (EXE), yolk sac and allantois) and intra-embryonic sites (embryonic ectoderm (EEC)). **c**, In the yolk sac, these progenitors aggregate into endothelial-lined blood islands that then fuse to generate a primary capillary plexus (see main text). **d**, The primary capillary plexus undergoes remodelling along with intra-embryonic vessels to form a mature circulation (**g**). **e**, Intra-embryonic angioblasts migrate along distinct pathways before (**f**) aggregating directly into the dorsal aorta or cardinal vein, without a plexus intermediate. **g**, The primary vessels (capillary plexus, dorsal aorta and cardinal vein) then remodel, together with the extra-embryonic plexus, to form a mature vasculature, which along with VEGF and Notch involves the angiopoietins and Tie receptors⁴. **h**, Mural cells (pericytes and smooth-muscle cells) proliferate and differentiate in response to transforming growth factor- β (TGF- β) signalling, and are recruited to vessels by platelet-derived growth factor (PDGF) secreted by endothelial cells^{5,6}. Ang, angiopoietin; Eph, Eph receptor family; Shh, sonic hedgehog; Np, neuropilin.

paracrine signals to surrounding cells, both in development and in ongoing endothelial niches in the adult. In addition, the signals that are primarily used to promote endothelial development might also act directly on other cell types, reinforcing the complex interplay between the vasculature and surrounding tissues. The second part of this review critically evaluates the evidence for the role of VEGF signalling in non-vascular tissues, as well as the role of the vasculature in inducing organogenesis and differentiation. Both of these processes have serious implications for the specificity of vascular therapies.

The haemangioblast/angioblast connection

In the mouse, the earliest marker of angioblast precursors is Flk1 (VEGFR-2), which is the major receptor for VEGF-A. Flk1 marks a subset of Brachyury-positive cells in the primitive streak, which then migrate into the extra-embryonic yolk sac to form a disperse vascular plexus, part of which contains clusters known as the blood islands (Fig. 2). The outer cells of the blood islands are endothelial, whereas the inner cells give rise to haematopoietic progenitors. In the embryo itself, angioblast precursors, marked by Flk1 expression, are dispersed throughout the head mesenchyme and other areas. The dorsal aorta and cardinal veins develop directly from aggregating angioblasts, whereas, in other areas, local vascular plexuses develop and slowly remodel and refine into the major vessels and capillary beds of the embryo. Flk1 is not only a marker for the earliest progenitors of the vascular and haematopoietic system, active VEGF-A signalling is also

required for normal development of both systems. Embryos lacking *Flk1* die at around 9 days of development and show no development of blood vessels or haematopoietic cells^{7,8}. Mutation of the ligand, VEGF-A, also leads to early lethality. Interestingly, VEGF-A is haploinsufficient — heterozygotes die early in gestation with severe reduction in the size and calibre of their developing blood vessels^{9,10}.

The loss of both vascular and haematopoietic cells in *Flk1* mutants is one of many pieces of information linking endothelial and haematopoietic development through a possible common progenitor, the haemangioblast (reviewed in ref. 11). Notably, Keller, Choi and colleagues showed that Flk1-positive cells sorted from differentiating embryonic stem (ES) cells can give rise to single-cell-derived blast colonies — so-called BL-CFCs — that can produce both endothelial and haematopoietic cells, thereby providing formal proof of the haemangioblast^{12,13}. In the embryo itself, not all endothelial progenitors have such bipotential fate. In the chick, only a subset of endothelial progenitors have haematopoietic potential. In the mouse, there are clear Flk1-positive progenitors in regions of the head mesenchyme and in non-blood-island regions of the yolk sac that do not have haematogenic potential¹¹. However, Flk1-positive cells with bipotential activity can be isolated from the primitive streak in greater numbers than from the early yolk sac¹⁴. This raises the possibility that all early Flk1-positive cells have the potential to become either haematopoietic or endothelial cells.

Isolation of single Flk1-positive cells that can give rise to both endothelial and haematopoietic cells *in vitro* is clear proof of a com-

mon progenitor for the two lineages. However, it does not exclude the possibility that the isolated precursor is a more primitive multipotent precursor, the potency of which has not been revealed by the assays used. During embryonic stem (ES) cell differentiation, isolated Flk1-positive progenitors can produce smooth-muscle-type cells^{15,16} or cardiomyocytes *in vitro*¹⁷ under appropriate culture conditions. These results suggest a broader potential for Flk1-positive progenitors than simply haematopoiesis and vasculogenesis, and imply that the Flk1-positive cells isolated from the primitive streak and from ES cells *in vitro* represent a multipotent mesodermal progenitor that later specializes into different mesodermal lineages. Whether endothelial progenitors in later embryonic or adult development have multipotent capacity is not known; however, this could help to explain recent evidence linking *Flk1* expression with multipotent stem-cell populations, such as MAP-C cells isolated from adult bone marrow¹⁸ and multipotent mesodermal (mesoangioblast) cell lines derived from developing dorsal aorta¹⁹.

Other studies have challenged the concept of a bipotential haemangioblast. Cell-lineage tracing of cells from the primitive streak to the yolk sac failed to reveal a common haematopoietic and endothelial progenitor²⁰, and other studies have suggested that haematopoietic and endothelial progenitors can be distinguished by their differential expression of CD41 as soon as they exit from the primitive streak²¹. What might seem contradictory results could really reflect issues of timing of the progression of lineage commitment from the primitive streak. Recent advances in imaging technologies²² might finally allow us to follow the fate of Flk1 progenitors in the living embryo and resolve these controversies.

From endothelial progenitors to a functioning circulation

Once the vascular progenitors have been specified, they begin to form a disperse vascular plexus, which is then gradually reorganized into a functional circulation. As vessels begin to be remodelled, they undergo localized proliferation and regression, as well as programmed branching and migration into different regions of the body. They need to be specified into different calibres and types of vessel, including division into arteries, veins and lymphatics, with further subdivision into large vessels, venules, arterioles, capillaries and so on. In addition, they need to recruit supporting cells, smooth-muscle cells and pericytes, to ensure the stability of the vessels formed (Fig. 2). Although we do not fully understand the intricacies of these processes, it is clear that the final outcome is determined by a combination of hard-wired genetic programming and extrinsic influences, such as hypoxia²³ and haemodynamic flow²⁴.

One of the first events that takes place during maturation of the circulatory system is the specification of arteries versus veins, and this serves as a good example of the interplay of intrinsic and extrinsic factors. Until fairly recently, it was assumed that specification of arterial versus venous fate was a late event in development and was instigated by haemodynamic-flow differences in the two types of vessel. This view changed when specific markers of venous versus arterial fate, such as ephrinB2 for arteries and its receptor EphB4 or veins^{25–27}, were detected on subsets of developing blood vessels before the onset of circulation. In the chick, the early extra-embryonic blood islands contain a random mixture of subpopulations of cells expressing neuropilin 1 (NP1), which is the VEGF co-receptor later restricted to arteries, and NP2, which is the vein-specific receptor. By the 13-somite stage, the expression of the two genes is segregated to the future arterial and venous parts of the plexus, despite the absence of blood flow²⁸. This study raises the possibility that arterial versus venous fate is established in early progenitors, which then segregate; a similar conclusion was suggested by cell-lineage tracing in zebrafish²⁹.

Although the pendulum has swung towards intrinsic specification of arterial versus venous fate, there is still considerable plasticity in the early vasculature. Grafting experiments in the chick have shown that, up to embryonic day 7, ectopic grafts of individual arteries or veins can lead to respecification of cell fate in early stages, but plasticity is grad-

ually lost later in development³⁰. Elegant time-lapse cinematography in the chick and zebrafish has shown that the establishment of the final pattern of circulation between artery and vein involves selective disconnection and reconnection of small vessels, and concomitant switches in markers of arterial versus venous fate^{24,31}. Manipulating flow physically (chick) or genetically (fish) showed that it could change gene expression and cell fate. Therefore, the overall vascular architecture is probably refined by the haemodynamics of circulatory-flow patterns on top of an underlying genetic programming of arterial versus venous fate^{24,29}.

The Notch signalling pathway has been implicated as a prime player initially, from studies in zebrafish, in establishing arterial versus venous fate. Injection of a dominant-negative form of Suppressor of Hairless (the common downstream effector of Notch signalling) resulted in decreased expression of arterial markers, such as ephrinB2, and ectopic expression of venous markers in the dorsal aorta^{29,32}. Conversely, expression of an activated Notch construct induced ectopic arterial markers in the posterior cardinal vein³². This suggests that during normal development, activation of the Notch signalling pathway in developing artery cells is required to suppress the venous fate and allow arterial differentiation. The hairy-related basic helix–loop–helix transcriptional repressors of the Hes/Hey family are direct downstream targets of Notch signalling in many situations in which Notch determines cell-fate choice. The zebrafish gridlock gene³³ encodes a member of this family of proteins. Gridlock is expressed specifically in the aorta, not the developing veins, which is consistent with its being a Notch target *in vivo*³³. Graded reduction in gridlock expression leads to progressive loss of the artery and expansion of the adjacent vein, in a similar manner to the effect seen by blocking Notch signalling with a dominant-negative Su(H) construct²⁹.

Four different Notch receptors (Notch 1–4) and five ligands (delta-like (Dll)1, Dll3, Dll4, Jagged-1 (Jag1) and Jag2) have been identified in mammals³⁴. Genetic analysis in mice and humans has revealed various types of vascular defect associated with Notch-pathway mutants (Supplementary Table 1), although defective primary separation of arterial versus venous fate seems to be a consistent feature of these complex phenotypes. However, mutations in Dll4, which is a Notch ligand specifically expressed in developing arterial but not venous endothelium, lead to defective development of the dorsal aorta and cardinal veins, with development of arterio-venous shunts^{35,36}. This is associated with downregulation of arterial markers and upregulation of venous markers in the dorsal aorta. Dll4 seems to act primarily in an autocrine manner on the arterial endothelium, through the receptors Notch1 and Notch4. Double mutants of Notch1 and Notch4 have a similar phenotype to loss of Dll4, and endothelial-specific knockout of RBP, the Su(H) orthologue, also leads to complex vascular defects, including loss of arterial specification³⁷. Double mutation of Hes1 and Hey1, mammalian orthologues of gridlock, also produces loss of arterial markers and vascular shunts³⁸. Thus, activation of Notch signalling and its downstream-response genes seems to be a conserved requirement for specification of arterial-cell fate in vertebrates, with venous fate being the 'default' state. Recent evidence has shown that the orphan nuclear receptor COUP-TFII is expressed specifically in venous endothelium and that mutation leads to activation of arterial markers in veins³⁹. Expression of COUP-TFII ectopically in arteries suppresses the expression of NP1 and other arterial markers. Together, these results suggest that COUP-TFII is a part of an active pathway promoting venous fate and suppressing Notch signalling.

Patterning the developing vasculature along the body axis

Although levels of Notch signalling might be crucial to establishing arterial versus venous fate, there must be other mechanisms that establish the positions in which the various primary blood vessels develop. The formation of the dorsal aorta and cardinal veins in chicks, and presumably in mice, occurs by local aggregation of angioblast progenitors in the midline and recruitment by inward migration of haematogenic progenitors to the ventral side of the dorsal aorta (Fig. 3). In zebrafish,

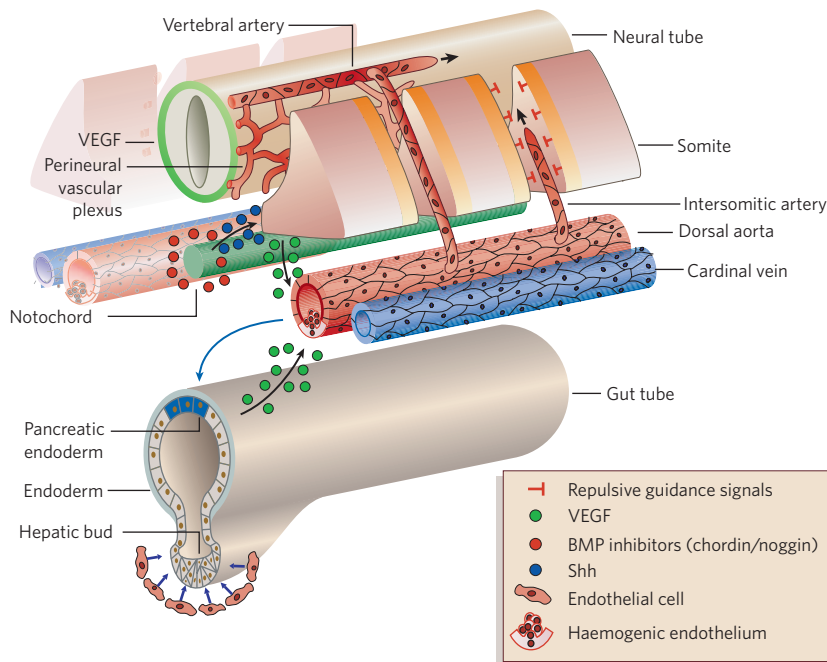


Figure 3 | Vascular development along the mammalian body axis. Formation of the dorsal aorta and cardinal veins occurs by *de novo* aggregation of angioblasts along the anterior/posterior body axis in response to VEGF from the endoderm. Somites, in response to hedgehog from the notochord, might also produce VEGF required for development of these axial vessels. Bone morphogenetic protein (BMP) inhibitors, noggin and chordin, produced by the notochord establish an avascular midline zone. Intersomitic vessels sprout from the dorsal aorta and cardinal vein (not shown) in a VEGF-dependent manner, and are then guided between the somites by repulsive cues, such as ephrinB2, semaphorin3A (sema3A) and sema3E in the somite boundary. Intersomitic arteries bifurcate once they reach the medial edge of the somite, fusing with adjacent intersomitic arteries to form the vertebral artery. The perineural vascular plexus surrounds the neural tube and is formed by angioblasts migrating from the somites in response to VEGF from the neural tube (green band). Once specified by VEGF from the endoderm, the dorsal aorta signals back to the endoderm to induce pancreatic bud formation by means of an unknown signal (blue arrow). Endothelial cells surrounding the prospective hepatic endoderm also provide an unknown signal (blue arrows) to induce hepatic bud development.

the single dorsal aorta might be induced by signals from the notochord, as *floating-head* and *no-tail* mutants lacking a notochord also lack a dorsal aorta. Ectopic transplantation of notochord cells allows assembly of aortic primordia^{40,41}. In *Xenopus*, the hypochord, which is a structure ventral to and induced by the notochord, is required for dorsal aorta formation^{42,43}. There is no hypochord in mice or chicks, but the definitive endoderm ventral to the aorta might take over the role of the hypochord in inducing dorsal aorta formation in these organisms. The common signal in all cases again seems to be VEGF-A, which acts in this context as a graded attractant to angioblasts to promote aorta formation. The hypochord in frogs, and definitive endoderm in chicks and mice, are potent sources of VEGF^{42,44}, whereas VEGF produced by the somites in response to sonic hedgehog (Shh) from the notochord promotes recruitment of cells to form the midline vessels in zebrafish⁴⁵. Hedgehog signalling also seems to be important in dorsal aorta formation in chicks and mice, as mutants for the hedgehog signalling component *smoothened*, or use of *smoothened* inhibitors, cause defective dorsal aorta formation⁴⁶. The nature of the hedgehog signal is less clear, as loss of Shh does not lead to loss of the dorsal aorta in mice, suggesting the involvement of other hedgehog ligands. The role of the notochord in dorsal aorta formation in chicks and mice is also not clear. *Brachyury* mutants lack the posterior notochord, but still form a posterior dorsal aorta⁴⁷. However, this does not rule out the role of the notochord in the anterior region, where the *smoothened* phenotype is most severe⁴⁶.

The notochord does have at least one key role in patterning the axial vasculature in chicks and mice by secreting bone morphogenetic protein (BMP) inhibitors, noggin and chordin, which define an avascular region around the developing notochord^{48,49}. Later, secretion of VEGF from the neural tube is involved in recruiting somite-derived angioblasts to form the perineural vascular plexus, which will encase the neural tube at midgestation⁵⁰. From there, angiogenic sprouting within the neural tube will occur and is dependent upon Shh signals from the ventral neural tube and neighbouring tissues⁵¹.

The formation of the intersomitic arteries from the dorsal aorta is a highly stereotyped process that begins at an early somite stage and spreads caudally as somites develop. The intersomitic vessels sprout dorsally from the dorsal aorta in the region between somites (Fig. 3). Once they reach the medial surface they are deflected along the surface of the somite, where they fuse with adjacent sprouts and form the

longitudinal vertebral arteries aligning with the neural tube. Beginning at the eight-somite stage, presumptive cardinal vein angioblasts send processes to join the vertebral artery and become the intersomitic veins dorsal to the intersomitic arteries⁵². What controls the position and timing of sprouting and migration of the intersomitic vessels? Again, VEGF has a role. Mutation of Flt1 (VEGFR-1), which is normally considered a negative regulator of VEGF signalling, reduced sprouting in ES cells *in vitro*, and intersomitic sprouts *in vivo* suggest that precise levels of bioactive VEGF-A and perhaps spatial localization of the VEGF signal are likely to determine proper localized intersomitic sprout formation⁵³. Notch signalling might also be important in determining the basic patterning and number of branch points, as well as in artery-vein specification. Several *in vitro* studies show a role for active Notch signalling in restricting branching morphogenesis and *in vivo* mutation studies support this. Loss of Dll4 (ref. 35) or Notch1 (ref. 54) leads to excessive and misdirected intersomitic branching, while activation of Notch signalling in the endothelium suppresses branching of vessels.

Once formed, intersomitic sprouts need to be properly guided in their migration between adjacent somites. This process, like many other examples of vessel formation and migration, has several mechanistic and genetic similarities to axon guidance. The roles of axon-guidance cues in blood-vessel development have recently been reviewed elsewhere⁵⁵.

Clearly, even the earliest phases of vessel patterning involve a complex hierarchy of signals from VEGF through Notch to specific vessel-guidance mechanisms using molecular strategies first implemented in the nervous system. These pathways also seem to be reactivated in the adult in situations of neoangiogenesis, and therefore become interesting new targets for pro-angiogenic or antiangiogenic therapies (see review in this issue by Ferrara and Kerbel, p. 967). The formation of the developing vasculature depends on tight regulation of developmental cues with some signalling pathways, such as VEGF and Dll4, showing haploinsufficient defects. If this sensitivity is manifested in situations of adult neoangiogenesis, then there is real hope for effective therapeutic interventions.

Endothelial cells and vascular signalling in organogenesis

The crucial pro-angiogenic properties of VEGF, as described above, have made it a prime candidate for therapeutic intervention involving

angiogenesis. The intimate association of endothelial cells with their cognate organs throughout life, and the potentially pleiotropic nature of VEGF signalling, might, however, lead to significant off-target effects from these interventions. In shaping the vascular system, organs signal to the vessels that service them, influencing the endothelial cells to adopt functional specialties, such as the blood–brain barrier and fenestrated endothelium in the kidney glomeruli⁵⁶. However, there is increasing evidence that endothelial cells, in turn, provide instructive morphogenetic cues to promote organ formation and patterning both in development and in the adult.

Endothelial cells, and liver and pancreas induction

The liver first appears as a multilayered epithelium in the ventral foregut endoderm. Endothelial cells surround this presumptive liver bud, then invade it and aggregate into sinusoids as hepatoblasts begin to migrate from the endoderm into the underlying septum transversum⁵⁷ (Fig. 3). A role for endothelial cells in liver development was suggested by studies of *Flk1*-mutant mice, which fail to form mature

endothelial cells⁷. In these mice, thickening of the foregut endoderm and expression of liver-specific genes occurred normally. However, there was no migration or proliferation of the hepatoblasts into the surrounding septum transversum to form the liver bud⁵⁷. Explant cultures demonstrated that the effect of endothelial cells on liver development was independent of circulation⁵⁷. Endothelial cells activated by VEGF also support hepatocyte proliferation and survival in adult animals⁵⁸. In response to VEGF, liver sinusoidal endothelial cells (LSEC) released hepatic mitogens, hepatic growth factor (HGF) and interleukin-6 (IL-6), in a VEGFR-1-dependent manner, promoting hepatic growth and protecting hepatocytes from toxic insult *in vivo*. Therefore, endothelial cells provide essential cues to the developing liver and can be stimulated to provide both nutritional and trophic support to a damaged adult liver.

Endothelial cells seem to have a similar role in promoting early pancreatic development. The developing pancreas forms in close association with the dorsal aorta and vitelline veins (Fig. 3)⁵⁹. Pancreatic islet endocrine cells (such as insulin-producing cells) also associate closely with endothelial cells, secreting hormones directly into the blood. Reciprocal signalling seems to occur between endothelial and endocrine cells in order to establish a functional pancreas. Explant studies showed that the dorsal aorta was necessary and sufficient for insulin expression in endoderm tissue, and ectopic endothelial cells could induce insulin expression in non-pancreatic endoderm⁵⁹. Endocrine cells then signal back to endothelial cells through VEGF to induce capillary invasion of the islets and fenestration of endothelial cells occupying the islets⁶⁰. VEGF was not required for endocrine pancreas development itself⁶⁰. Analysis of *Flk1*-mutant mice suggested that endothelial cells were not required for initial specification of the pancreas from foregut endoderm but were required for the subsequent emergence of the pancreatic bud and expression of Ptf1a, which is a transcription factor necessary for pancreatic development, as well as endocrine genes⁶¹.

Studies in zebrafish have not supported this early role for endothelial cells in liver and pancreas development, as liver and pancreatic budding occurs normally in zebrafish mutants lacking endothelial cells^{62,63}. This might reflect evolutionary differences in the source of morphogenetic signals to compensate for differences in development and function of these organs between fish and mammals⁶⁴. The signals provided by the endothelium to promote liver and pancreatic bud morphogenesis are currently unknown. Determining the nature of these signals, their conservation between fish and mammals, and whether they are expressed by all endothelial cells or only those in close juxtaposition to the liver and pancreas will provide interesting areas for future research.

Endothelial cells, VEGF and the kidney glomerulus

Numerous studies in mice have shown that signalling between endothelial cells and podocytes is essential for proper development and maintenance of the filtration function of the kidney glomerulus^{65,66}. Podocytes, which are the specialized cells that make up the support structures of the functional glomerulus, normally express VEGF at high levels. Homozygous deletion of VEGF specifically in podocytes prevented recruitment and maturation of endothelial cells in the glomerulus and led to abnormal podocyte and mesangial-cell maturation and perinatal lethality⁶⁵. Mesangial-cell contribution to glomeruli was also abnormal in mice specifically lacking PDGFB in endothelial cells⁶⁷. Together, these results suggest that endothelial cells, which are recruited, matured and maintained by VEGF from podocytes, promote further maturation of podocytes and mesangial cells, and formation of a functional glomerulus. Endothelial-cell maintenance through regulated VEGF levels is crucial for continued glomerular function in adults. Heterozygous loss of VEGF in podocytes resulted in the disappearance of endothelial fenestrations in adult mice, followed by loss of podocyte foot processes. Clinical symptoms of hypertension and proteinuria were followed by end-stage kidney failure⁶⁵. Administration of VEGF-neutralizing agents results

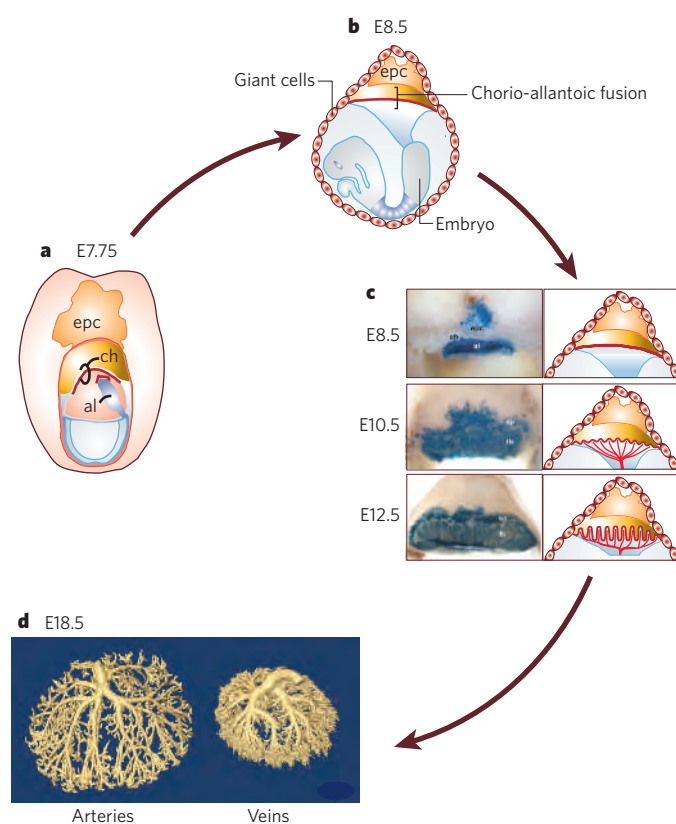


Figure 4 | Formation of the fetal vasculature in the chorio-allantoic placenta.

a, Fetal components of the chorio-allantoic placenta as shown at embryonic day (E) 7.75 are the chorionic membrane (ch) and allantois (al). Future fusion areas are highlighted in red. **b**, At the four-to-six-somite stage, the allantois fuses to the chorion, forming the chorio-allantoic placenta. **c**, Shortly after fusion, the chorion branches, forming the labyrinth layer that is invaded by fetal vasculature from the allantois. The panels on the left show whole-mount staining for Flt1, revealing a subset of ectoplacental cone cells, spongiotrophoblasts, and trophoblasts and endothelial cells in the labyrinth layer. (Image reproduced with permission from ref. 73.) Vessels invading the labyrinth layer undergo extensive remodelling to form a complex vascular structure of fetal arteries and veins. **d**, Surface renderings derived from reconstructed micro-computed tomography showing placental vasculature at E18.5, displaying the intricate nature of the fetal vasculature of the placenta. Contrast agent was injected separately into the placental arteries or veins via the umbilical vessel vasculature. (Image courtesy of M. Y. Rennie, S. L. Adamson and J. G. Sled of the Mouse Imaging Centre, Hospital for Sick Children, Toronto, Canada, and Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.)

in glomerular dysfunction, suggesting that the glomerulus is highly dependent on endothelial function controlled by precise VEGF levels throughout life⁶⁸.

Endothelial cells and placental development

The mammalian placenta is a complex organ, the main functions of which depend on the establishment of a structure in which oxygen, waste products, nutrients and growth factors can be interchanged between the maternal and fetal environments. Establishment of the vascular-interchange bed on the fetal side involves fusion of the allantois to the chorion and branching invasion of the fetal capillaries into the chorionic trophoblast (Fig. 4). On the maternal side, the developing spiral arteries entering the placenta undergo a process of invasion by fetal trophoblasts, which eventually replace the endothelial cells, ensuring that maternal blood directly bathes fetal trophoblasts and thereby enhancing exchange.

Despite some detailed differences in architecture, the overall gross anatomical features and molecular pathways underlying placental development are similar in humans and mice⁶⁹. Mouse mutants in many pathways affecting vascular development cause placental problems⁶⁹ and the same pathways are probably relevant to human placental development. However, on the whole, the pathways involved in placental vascular development are similar to those involved in the establishment of other vascular beds. A few special features do arise, which might be related to the rapid and extensive neovascularization and remodelling that the placenta needs. First, the placenta secretes a number of specific signalling molecules that seem to be involved in promoting vascular development, beyond the standard angiogenic factors. For example, the placental lactogen-related hormones, proliferin and proliferin-related protein, exert angiogenic and anti-angiogenic actions on the placental vasculature⁷⁰. In addition, one member of the VEGF family, placental-like growth factor (PlGF), is strongly expressed in the placenta and is thought to play its normal physiological role there. Although mice lacking PlGF show no apparent physiological defects in normal development⁷¹, they do show defects in situations of induced angiogenesis, such as tissue ischaemia. Binding of PlGF to Flt1 enhances VEGF signalling through Flk1 (ref. 72), suggesting a mechanism by which this molecule might have a modulatory role in the developing placenta.

Interestingly, in both mice and humans, Flt1 (the VEGFR that preferentially binds PlGF) is expressed not only in the fetal vasculature but also on subsets of trophoblast cells within the placenta, suggesting that PlGF also has a direct role in promoting placental development. However, chimaeric mouse placentas, in which trophoblast cells lack Flt1 but fetal vascular expression is intact, are morphologically normal and support fetal growth to term⁷³. Therefore, trophoblast-specific expression of Flt1 is dispensable for establishment of the maternal–fetal interface and the formation of placental vasculature. This leaves open the possibility that the placental expression of PlGF and Flt1 could still have an important modulatory role in the placenta itself or, by virtue of the production of the soluble form of Flt1, sFlt1, on the maternal system.

Pre-eclampsia is a relatively common disorder affecting about 2.5–5% of pregnancies (reviewed in ref. 74). It appears in two stages: the preclinical phase is characterized by failure of cytotrophoblast invasion of maternal spiral arterioles, leading to a hypoxic placenta and, hence, upregulation of the production of hypoxic-responsive placental factors secreted by the trophoblast, including components of the VEGF/PlGF pathway. In turn, this local defect leads to the second phase of pre-eclampsia whereby systemic responses in the mother result in hypertension, proteinuria, blood clotting and various other internal organ dysfunctions predominantly due to endothelial dysfunction. Pre-eclampsia can be life threatening to both mother and baby and can be cured only by delivery of the placenta. Recent evidence has suggested that increased circulating levels of sFlt1, along with reduced levels of VEGF and placental growth factor (PlGF) in maternal serum 5 weeks before clinical onset of pre-eclampsia, can be

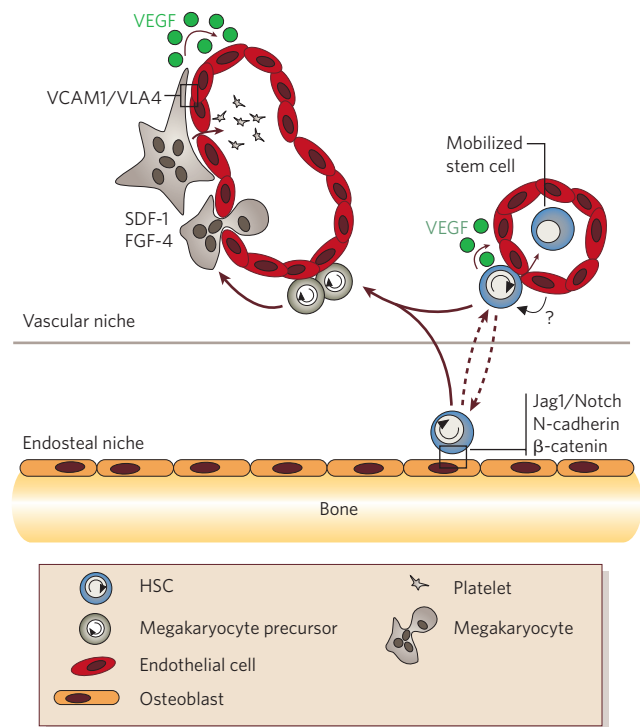


Figure 5 | Endothelial cells provide a niche for haematopoietic stem cells (HSCs).

Osteoblasts in the endosteal niche expressing Jagged-1 (Jag1) and N-cadherin contact and maintain HSCs by activation of Notch, and might further regulate HSC activity through N-cadherin and β -catenin signalling. Endothelial cells in the vascular niche also contact HSCs and provide unknown maintenance signals (question mark). HSCs might be transported between niches and could be subject to differential regulation in each niche (dashed lines). Endothelial cells expressing vascular cell-adhesion molecule 1 (VCAM-1) also associate closely with megakaryocytes and their progenitors through very late activation antigen 4 (VLA4) in response to chemotactic factors, stromal cell-derived factor-1 (SDF1) and fibroblast growth factor-4 (FGF4), and provide a niche for megakaryocyte maturation and platelet production. The immediate juxtaposition of HSCs to endothelial cells also facilitates their rapid mobilization and entry into circulation in response to stress and, in the case of megakaryocytes, release of platelets directly into the blood. HSCs and haematopoietic progenitor cells as well as megakaryocytes produce VEGF and other angiogenic factors, which might act in a feedback loop to support endothelial cells in the bone marrow and in the periphery at sites of normal and pathologic angiogenesis.

predictive of disease⁷⁵. Importantly, delivery of sFlt1 directly to pregnant rats recapitulated many hallmarks of preeclampsia, including hypertension, proteinuria and kidney endotheliosis⁶⁸. Therefore, blocking sFlt1 is an attractive avenue to manage pre-eclampsia. The finding that Flt1 is not required for placental development itself in mice is encouraging, as it suggests that there might not be unwanted placental side effects of any such therapy.

Endothelial cells and the stem-cell niche

An additional function for endothelial cells in organogenesis might be the regulation of stem cells. This has been proposed for the nervous system. Proliferating endothelial and neural precursors are closely associated, and endothelial-derived soluble factors regulate neurogenesis (see review in this issue by Greenberg and Jin, p. 954). Given their close proximity, cross-talk between proliferating endothelial and neuronal progenitors is likely to involve more than just secreted factors. Juxtacrine signalling molecules, such as Notch, which are essential for cell-fate choice in progenitors of both neuronal and vascular lineages, are obvious candidates. Endothelial cells are therefore likely to have a key role in providing a niche for regulating neural stem-cell activity.

Endothelial cells are also found in other areas where stem cells reside and might have important roles in promoting their maintenance and survival. Haematopoietic stem cells (HSCs) are well defined in terms of their stem-cell properties, but the stromal environment contributing to their regulation is less well defined. One cell type contributing to the adult HSC niche is the osteoblast, but other stromal cells in the marrow, such as endothelial cells, probably also contribute^{76,77} (Fig. 5). As already discussed, there is an intimate association and, indeed, close lineage relationship of endothelial and blood cells, suggesting that they also regulate HSC development from their inception through to their occupation of the bone marrow. Bone-marrow endothelial cells regulate proliferation and differentiation of more-committed progenitors of the myeloid and megakaryocyte lineages^{78,79}. As megakaryocytes produce VEGF, the interaction between these progenitors and endothelial cells is probably reciprocal, as it is in other organs⁸⁰. Endothelial cells also promote survival of HSCs in culture, but this seems to be limited to certain populations of endothelial cells⁸¹. On the basis of this finding, the existence of an endothelial niche for HSCs seems axiomatic; however, only recently has *in vivo* evidence emerged supporting an endothelial niche for HSCs⁸². Significant fractions of HSCs in both adult bone marrow and spleen were found in close association with endothelial sinusoids, suggesting that endothelial cells provide support to HSCs *in vivo*. It will be interesting to further define the respective contributions of endothelial and endosteal niches to HSC behaviour. Furthering the idea of reciprocal signalling between haematopoietic progenitors and endothelial cells, bone-marrow-derived progenitor cells, possibly including haematopoietic progenitor cells and HSCs, localize to sites of active angiogenesis, including tumour angiogenesis, raising the possibility that haematopoietic cells signal back to endothelial cells to regulate angiogenesis⁸³. The transport of circulating HSCs or haematopoietic progenitor cells to specific organs might be dependent on adhesion molecules specific to endothelial cells of each organ and could represent another level of regulation.

Further study of the functional roles of endothelial cells in promoting adult organ maintenance, and stem-cell and progenitor-cell proliferation, seems certain to reveal more interesting functions for the vasculature than simply carrying the blood supply.

VEGF-A signalling in non-endothelial cells

Expression of VEGFRs on non-endothelial cells and observations made in VEGF-mutant mice have also implicated VEGF as an essential signalling factor in non-endothelial cells. However, the evidence for a direct effect versus indirect effects needs to be critically evaluated in all cases.

VEGF action on cells of the nervous system

Numerous studies point to a neuroprotective role for VEGF⁸⁴. For instance, mice homozygous for a VEGF-A allele lacking the hypoxia-response element developed a motor neuron disease similar to amyotrophic lateral sclerosis (ALS)⁸⁵. This effect of reduced VEGF was probably partly due to effects on endothelial cells, as reduced neural perfusion was observed; however, a direct effect of VEGF might also be involved, as VEGF protected motor neurons under stress conditions *in vitro*⁸⁵ and overexpression of Flk1 in motor neurons was subsequently shown to prolong survival of ALS mice⁸⁶. Indeed, *in vitro*, VEGF displays pro-survival activity for many types of neuron under a range of stress conditions, stimulates axon outgrowth in explant cultures, and promotes the survival, proliferation and migration of Schwann cells, astrocytes and microglia⁸⁴. *In vivo* studies have demonstrated that VEGF-1 (ref. 64) is required for migration of facial motor neurons⁸⁷. Lack of VEGF-A specifically in the neural population resulted in abnormalities in retina and cortex development, accompanied by reduced vascular density and branching in these sites, which, in the retina, resembled some human retinopathies⁸⁸. However, neural development proceeded apparently normally when Flk1 was removed from the neural population, sug-

gesting that the VEGF-A phenotype was caused by secondary effects of loss of appropriate vascular invasion⁸⁸. VEGF might therefore have two roles in nervous-system development and neuroprotection: first, as a direct neurotrophic agent; and second, as an angiogenic factor stimulating endothelial cells to provide adequate perfusion and neurotrophic factors. Regardless of the direct or indirect role, administration of VEGF can reduce the severity of neurological trauma, including spinal-cord injuries, making it a potentially useful adjunct therapy for these patients⁸⁹.

Bone

Signalling by VEGF seems to have multiple roles in bone development, promoting vascularization during endochondral bone formation, and regulating survival and activity of chondrogenic and osteogenic cells. Soluble inhibitors first revealed a requirement for VEGF in bone development, suppressing vascular invasion and growth of long bones⁹⁰. Conditional and isoform-specific VEGF mutants have subsequently shown that the heparin-binding isoforms of VEGF produced by chondrocytes are essential for vascular invasion of the metaphysis, cartilage resorption and primary ossification of long bones^{91,92}. VEGF, presumably secreted from non-hypertrophic chondrocytes, also seems to be required to promote angiogenesis around the epiphysis of long bones to facilitate secondary ossification. In contrast to metaphyseal vasculature, recruitment of epiphyseal vasculature depends on the soluble VEGF isoforms⁹³, the absence of which results in massive apoptosis of non-hypertrophic epiphyseal chondrocytes due to hypoxia^{93,94}. VEGF might act directly on these chondrocytes to protect them against hypoxia^{93,94}. VEGF also has an important role in bone-fracture healing in animal models, in part by stimulation of vascular growth in the region of the injury, but evidence also suggests that it acts on bone-forming cells themselves⁹⁵. VEGF directly stimulated the activity of isolated osteoblasts and osteoclasts *in vitro*, whereas VEGF blockade in explant cultures of embryonic cartilaginous metatarsals inhibited ossification^{95,96}.

VEGF might signal to other non-endothelial cell types in addition to those described. For example, VEGF deletion in HSCs suggests an autocrine VEGF signalling loop in HSC survival⁹⁷, whereas a podocyte-specific *Flk1* mutation points to an autocrine role for VEGF in adult podocytes (S. Quaggin, personal communication). Other non-endothelial targets of VEGF might include myoblasts⁹⁸ and pneumocytes⁹⁹. The examples presented above clearly highlight the need for more studies using cell-specific VEGFR mutants to further dissect the autonomous and non-autonomous requirements of VEGF signalling in both development and disease.

Future directions

We have presented an overview of vascular development from the initial specification of vascular cell types through to their role in maintaining homeostasis in adult tissues. Although we have made significant progress towards understanding how blood vessels are formed and patterned to generate the many types of vessel, much work remains to be done. We still do not fully understand the origins of angioblasts or the relationship of endothelial cells to other vascular cell types, such as cardiac, smooth-muscle and blood cells. The complexity and heterogeneity of endothelial cells is becoming increasingly apparent, opening the possibility of refining therapeutic applications to specific subsets of the vasculature. To what extent is this heterogeneity defined by intrinsic programmes or influenced by local environment? The genetic and environmental control of vascular assembly and remodelling is still not well understood, although it is crucial to understanding the states of neoangiogenesis in the adult and defining new angiogenesis-based therapies. The multiple roles of VEGF signalling in endothelial development and function have made it the most popular target for angiogenic therapeutic interventions. Yet, its importance and the increasing evidence of its possible roles in cell types other than the vasculature make the side effects of anti-VEGF therapies a real concern. A more refined understanding of the complex sig-

nalling pathways controlling all aspects of vascular development and remodelling will help define new and more-specific targets for future therapeutic intervention. ■

- Carmeliet, P. Angiogenesis in health and disease. *Nature Med.* **9**, 653–660 (2003).
- Rossant, J. & Howard, L. Signaling pathways in vascular development. *Annu. Rev. Cell Dev. Biol.* **18**, 541–573 (2002).
- Bikfalvi, A. & Bicknell, R. Recent advances in angiogenesis, anti-angiogenesis and vascular targeting. *Trends Pharmacol. Sci.* **23**, 576–582 (2002).
- Thurston, G. Role of angiopoietins and Tie receptor tyrosine kinases in angiogenesis and lymphangiogenesis. *Cell Tissue Res.* **314**, 61–68 (2003).
- Betsholtz, C., Lindblom, P. & Gerhardt, H. Role of pericytes in vascular morphogenesis. *EXS* **115**–125 (2005).
- Lebrin, F., Deckers, M., Bertolino, P. & Ten Dijke, P. TGF- β receptor function in the endothelium. *Cardiovasc. Res.* **65**, 599–608 (2005).
- Shalaby, F. *et al.* Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–66 (1995).
- Shalaby, F. *et al.* A requirement for Flk1 in primitive and definitive hematopoiesis and vasculogenesis. *Cell* **89**, 981–990 (1997).
- Carmeliet, P. *et al.* Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* **380**, 435–439 (1996).
- Ferrara, N. *et al.* Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* **380**, 439–442 (1996).
- Ema, M. & Rossant, J. Cell fate decisions in early blood vessel formation. *Trends Cardiovasc. Med.* **13**, 254–259 (2003).
- Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725–732 (1998).
- Chung, Y. S. *et al.* Lineage analysis of the hemangioblast as defined by Flk1 and SCL expression. *Development* **129**, 5511–5520 (2002).
- Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. & Keller, G. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**, 625–630 (2004).
- Yamashita, J. *et al.* Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature* **408**, 92–96 (2000).
- Ema, M. *et al.* Combinatorial effects of Flk1 and Tal1 on vascular and hematopoietic development in the mouse. *Genes Dev.* **17**, 380–393 (2003).
- Schroeder, T. *et al.* Recombination signal sequence-binding protein Jg alters mesodermal cell fate decisions by suppressing cardiomyogenesis. *Proc. Natl Acad. Sci. USA* **100**, 4018–4023 (2003).
- Jiang, Y. *et al.* Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature* **418**, 41–49 (2002).
- Minasi, M. G. *et al.* The meso-angioblast: a multipotent, self-renewing cell that originates from the dorsal aorta and differentiates into most mesodermal tissues. *Development* **129**, 2773–2783 (2002).
- Kinder, S. J., Loebe, D. A. & Tam, P. P. Allocation and early differentiation of cardiovascular progenitors in the mouse embryo. *Trends Cardiovasc. Med.* **11**, 177–184 (2001).
- Ferkowicz, M. J. *et al.* CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. *Development* **130**, 4393–4403 (2003).
- Jones, E. A. *et al.* Dynamic *in vivo* imaging of postimplantation mammalian embryos using whole embryo culture. *Genesis* **34**, 228–235 (2002).
- Ramirez-Bergeron, D. L. *et al.* Hypoxia affects mesoderm and enhances hemangioblast specification during early development. *Development* **131**, 4623–4634 (2004).
- le Noble, F. *et al.* Flow regulates arterial-venous differentiation in the chick embryo yolk sac. *Development* **131**, 361–375 (2004).
- Adams, R. H. *et al.* Roles of ephrinB ligands and EphB receptors in cardiovascular development: demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis. *Genes Dev.* **13**, 295–306 (1999).
- Wang, H. U., Chen, Z. F. & Anderson, D. J. Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4. *Cell* **93**, 741–753 (1998).
- Gerety, S. S., Wang, H. U., Chen, Z. F. & Anderson, D. J. Symmetrical mutant phenotypes of the receptor EphB4 and its specific transmembrane ligand ephrin-B2 in cardiovascular development. *Mol. Cell* **4**, 403–414 (1999).
- Herzog, Y., Guttman-Raviv, N. & Neufeld, G. Segregation of arterial and venous markers in subpopulations of blood islands before vessel formation. *Dev. Dyn.* **232**, 1047–1055 (2005).
- Zhong, T. P., Childs, S., Leu, J. P. & Fishman, M. C. Gridlock signalling pathway fashions the first embryonic artery. *Nature* **414**, 216–220 (2001).
- Moyon, D., Pardanau, L., Yuan, L., Breant, C. & Eichmann, A. Plasticity of endothelial cells during arterial-venous differentiation in the avian embryo. *Development* **128**, 3359–3370 (2001).
- Isogai, S., Lawson, N. D., Torrealday, S., Horiguchi, M. & Weinstein, B. M. Angiogenic network formation in the developing vertebrate trunk. *Development* **130**, 5281–5290 (2003).
- Lawson, N. D. *et al.* Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development* **128**, 3675–3683 (2001).
- Zhong, T. P., Rosenberg, M., Mohideen, M. A., Weinstein, B. & Fishman, M. C. gridlock, an HLH gene required for assembly of the aorta in zebrafish. *Science* **287**, 1820–1824 (2000).
- Nye, J. S. & Kopan, R. Developmental signaling. Vertebrate ligands for Notch. *Curr. Biol.* **5**, 966–969 (1995).
- Duarte, A. *et al.* Dosage-sensitive requirement for mouse Dll4 in artery development. *Genes Dev.* **18**, 2474–2478 (2004).
- Gale, N. W. *et al.* Haploinsufficiency of delta-like 4 ligand results in embryonic lethality due to major defects in arterial and vascular development. *Proc. Natl Acad. Sci. USA* **101**, 15949–15954 (2004).
- Krebs, L. T. *et al.* Haploinsufficient lethality and formation of arteriovenous malformations in Notch pathway mutants. *Genes Dev.* **18**, 2469–2473 (2004).
- Fischer, A., Schumacher, N., Maier, M., Sendtner, M. & Gessler, M. The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. *Genes Dev.* **18**, 901–911 (2004).
- You, L. R. *et al.* Suppression of Notch signalling by the COUP-TFII transcription factor regulates vein identity. *Nature* **435**, 98–104 (2005).
- Fouquet, B., Weinstein, B. M., Serluca, F. C. & Fishman, M. C. Vessel patterning in the embryo of the zebrafish: guidance by notochord. *Dev. Biol.* **183**, 37–48 (1997).
- Sumoy, L., Keasey, J. B., Dittman, T. D. & Kimelman, D. A role for notochord in axial vascular development revealed by analysis of phenotype and the expression of VEGF-2 in zebrafish flh and ntl mutant embryos. *Mech. Dev.* **63**, 15–27 (1997).
- Cleaver, O. & Krieg, P. A. VEGF mediates angioblast migration during development of the dorsal aorta in *Xenopus*. *Development* **125**, 3905–3914 (1998).
- Cleaver, O. & Krieg, P. A. Notochord patterning of the endoderm. *Dev. Biol.* **234**, 1–12 (2001).
- Miquelot, L., Gertsenstein, M., Harpal, K., Rossant, J. & Nagy, A. Multiple developmental roles of VEGF suggested by a LacZ-tagged allele. *Dev. Biol.* **212**, 307–322 (1999).
- Liang, D. *et al.* The role of vascular endothelial growth factor (VEGF) in vasculogenesis, angiogenesis, and hematopoiesis in zebrafish development. *Mech. Dev.* **108**, 29–43 (2001).
- Vokes, S. A. *et al.* Hedgehog signaling is essential for endothelial tube formation during vasculogenesis. *Development* **131**, 4371–4380 (2004).
- Hogan, K. A. & Bautch, V. L. Blood vessel patterning at the embryonic midline. *Curr. Top. Dev. Biol.* **62**, 55–85 (2004).
- Reese, D. E., Hall, C. E. & Mikawa, T. Negative regulation of midline vascular development by the notochord. *Dev. Cell* **6**, 699–708 (2004).
- Nimmagadda, S. *et al.* BMP4 and noggin control embryonic blood vessel formation by antagonistic regulation of VEGFR-2 (Quek1) expression. *Dev. Biol.* **280**, 100–110 (2005).
- Hogan, K. A., Ambler, C. A., Chapman, D. L. & Bautch, V. L. The neural tube patterns vessels developmentally using the VEGF signaling pathway. *Development* **131**, 1503–1513 (2004).
- Nagase, T., Nagase, M., Yoshimura, K., Fujita, T. & Koshima, I. Angiogenesis within the developing mouse neural tube is dependent on sonic hedgehog signaling: possible roles of motor neurons. *Genes Cells* **10**, 595–604 (2005).
- Coffin, J. D. & Poole, T. J. Embryonic vascular development: immunohistochemical identification of the origin and subsequent morphogenesis of the major vessel primordia in quail embryos. *Development* **102**, 735–748 (1988).
- Kearney, J. B., Kappas, N. C., Ellerstrom, C., DiPaola, F. W. & Bautch, V. L. The VEGF receptor flt-1 (VEGFR-1) is a positive modulator of vascular sprout formation and branching morphogenesis. *Blood* **103**, 4527–4535 (2004).
- Krebs, L. T. *et al.* Notch signaling is essential for vascular morphogenesis in mice. *Genes Dev.* **14**, 1343–1352 (2000).
- Carmeliet, P. & Tessier-Lavigne, M. Common mechanisms of nerve and blood vessel wiring. *Nature* **436**, 193–200 (2005).
- Nikolova, G. & Lammert, E. Interdependent development of blood vessels and organs. *Cell Tissue Res.* **314**, 33–42 (2003).
- Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559–563 (2001).
- LeCouter, J. *et al.* Angiogenesis-independent endothelial protection of liver: role of VEGFR-1. *Science* **299**, 890–893 (2003).
- Lammert, E., Cleaver, O. & Melton, D. Induction of pancreatic differentiation by signals from blood vessels. *Science* **294**, 564–567 (2001).
- Lammert, E. *et al.* Role of VEGF-A in vascularization of pancreatic islets. *Curr. Biol.* **13**, 1070–1074 (2003).
- Yoshitomi, H. & Zaret, K. S. Endothelial cell interactions initiate dorsal pancreas development by selectively inducing the transcription factor Ptf1a. *Development* **131**, 807–817 (2004).
- Field, H. A., Ober, E. A., Roeser, T. & Stainier, D. Y. Formation of the digestive system in zebrafish. I. Liver morphogenesis. *Dev. Biol.* **253**, 279–290 (2003).
- Field, H. A., Dong, P. D., Beis, D. & Stainier, D. Y. Formation of the digestive system in zebrafish. II. Pancreas morphogenesis. *Dev. Biol.* **261**, 197–208 (2003).
- Lemaigre, F. & Zaret, K. S. Liver development update: new embryo models, cell lineage control, and morphogenesis. *Curr. Opin. Genet. Dev.* **14**, 582–590 (2004).
- Eremina, V. *et al.* Glomerular-specific alterations of VEGF-A expression lead to distinct congenital and acquired renal diseases. *J. Clin. Invest.* **111**, 707–716 (2003).
- Mattot, V. *et al.* Loss of the VEGF(164) and VEGF(188) isoforms impairs postnatal glomerular angiogenesis and renal arteriogenesis in mice. *J. Am. Soc. Nephrol.* **13**, 1548–1560 (2002).
- Bjarnegard, M. *et al.* Endothelium-specific ablation of PDGFB leads to pericyte loss and glomerular, cardiac and placental abnormalities. *Development* **131**, 1847–1857 (2004).
- Maynard, S. E. *et al.* Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. *J. Clin. Invest.* **111**, 649–658 (2003).
- Rossant, J. & Cross, J. C. Placental development: lessons from mouse mutants. *Nature Rev. Genet.* **2**, 538–548 (2001).
- Jackson, D., Volpert, O. V., Bouck, N. & Linzer, D. I. Stimulation and inhibition of angiogenesis by placental proliferin and proliferin-related protein. *Science* **266**, 1581–1584 (1994).
- Carmeliet, P. *et al.* Synergism between vascular endothelial growth factor and placental growth factor contributes to angiogenesis and plasma extravasation in pathological conditions. *Nature Med.* **7**, 575–583 (2001).
- Autiero, M. *et al.* Role of PlGF in the intra- and intermolecular cross talk between the VEGF receptors Flt1 and Flk1. *Nature Med.* **9**, 936–943 (2003).
- Hirashima, M., Lu, Y., Byers, L. & Rossant, J. Trophoblast expression of fms-like tyrosine kinase 1 is not required for the establishment of the maternal-fetal interface in the mouse placenta. *Proc. Natl Acad. Sci. USA* **100**, 15637–15642 (2003).
- Redman, C. W. & Sargent, I. L. Latest advances in understanding preeclampsia. *Science* **308**, 1592–1594 (2005).
- Levine, R. J. *et al.* Circulating angiogenic factors and the risk of preeclampsia. *N. Engl. J. Med.* **350**, 672–683 (2004).
- Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
- Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).

78. Avecilla, S. T. *et al.* Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis. *Nature Med.* **10**, 64–71 (2004).
79. Rafii, S. *et al.* Human bone marrow microvascular endothelial cells support long-term proliferation and differentiation of myeloid and megakaryocytic progenitors. *Blood* **86**, 3353–3363 (1995).
80. Mohle, R., Green, D., Moore, M. A., Nachman, R. L. & Rafii, S. Constitutive production and thrombin-induced release of vascular endothelial growth factor by human megakaryocytes and platelets. *Proc. Natl Acad. Sci. USA* **94**, 663–668 (1997).
81. Li, W., Johnson, S. A., Shelley, W. C. & Yoder, M. C. Hematopoietic stem cell repopulating ability can be maintained *in vitro* by some primary endothelial cells. *Exp. Hematol.* **32**, 1226–1237 (2004).
82. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
83. Rafii, S. & Lyden, D. Therapeutic stem and progenitor cell transplantation for organ vascularization and regeneration. *Nature Med.* **9**, 702–712 (2003).
84. Störkebaum, E. & Carmeliet, P. VEGF: a critical player in neurodegeneration. *J. Clin. Invest.* **113**, 14–18 (2004).
85. Oosthuysen, B. *et al.* Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration. *Nature Genet.* **28**, 131–138 (2001).
86. Störkebaum, E. *et al.* Treatment of motoneuron degeneration by intracerebroventricular delivery of VEGF in a rat model of ALS. *Nature Neurosci.* **8**, 85–92 (2005).
87. Schwarz, Q. *et al.* Vascular endothelial growth factor controls neuronal migration and cooperates with Sema3A to pattern distinct compartments of the facial nerve. *Genes Dev.* **18**, 2822–2834 (2004).
88. Haigh, J. J. *et al.* Cortical and retinal defects caused by dosage-dependent reductions in VEGF-A paracrine signaling. *Dev. Biol.* **262**, 225–241 (2003).
89. Rosenstein, J. M. & Krum, J. M. New roles for VEGF in nervous tissue—beyond blood vessels. *Exp. Neurol.* **187**, 246–253 (2004).
90. Gerber, H. P. *et al.* VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. *Nature Med.* **5**, 623–628 (1999).
91. Maes, C. *et al.* Impaired angiogenesis and endochondral bone formation in mice lacking the vascular endothelial growth factor isoforms VEGF164 and VEGF188. *Mech. Dev.* **111**, 61–73 (2002).
92. Zelzer, E. *et al.* Skeletal defects in VEGF(120/120) mice reveal multiple roles for VEGF in skeletogenesis. *Development* **129**, 1893–1904 (2002).
93. Maes, C. *et al.* Soluble VEGF isoforms are essential for establishing epiphyseal vascularization and regulating chondrocyte development and survival. *J. Clin. Invest.* **113**, 188–199 (2004).
94. Zelzer, E. *et al.* VEGFA is necessary for chondrocyte survival during bone development. *Development* **131**, 2161–2171 (2004).
95. Street, J. *et al.* Vascular endothelial growth factor stimulates bone repair by promoting angiogenesis and bone turnover. *Proc. Natl Acad. Sci. USA* **99**, 9656–9661 (2002).
96. Nakagawa, M. *et al.* Vascular endothelial growth factor (VEGF) directly enhances osteoclastic bone resorption and survival of mature osteoclasts. *FEBS Lett.* **473**, 161–164 (2000).
97. Gerber, H. P. *et al.* VEGF regulates haematopoietic stem cell survival by an internal autocrine loop mechanism. *Nature* **417**, 954–958 (2002).
98. Arsic, N. *et al.* Vascular endothelial growth factor stimulates skeletal muscle regeneration *in vivo*. *Mol. Ther.* **10**, 844–854 (2004).
99. Compernelle, V. *et al.* Loss of HIF-24 and inhibition of VEGF impair fetal lung maturation, whereas treatment with VEGF prevents fatal respiratory distress in premature mice. *Nature Med.* **8**, 702–710 (2002).
100. Sharpe, J. *et al.* Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science* **296**, 541–545 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Quaggin (Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada) for unpublished results, J. Walls (Mouse Imaging Centre, Hospital for Sick Children, Toronto, Canada) for assistance with preparation of Fig. 1, and M. Hirashima (Keio University, Tokyo, Japan) and members of the Rossant laboratory for helpful discussions. J.R. is a distinguished investigator of the Canadian Institutes of Health Research. L.C. is a postdoctoral fellow of the Canadian Institutes of Health Research and K.C. is a C. J. Martin Fellow of the Australian National Health and Medical Council.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to J.R. (e-mail: janet.rossant@sickkids.ca).

Lymphangiogenesis in development and human disease

Kari Alitalo¹, Tuomas Tammela¹ & Tatiana V. Petrova²

The lymphatic vasculature forms a vessel network that drains interstitial fluid from tissues and returns it to the blood. Lymphatic vessels are also an essential part of the body's immune defence. They have an important role in the pathogenesis of several diseases, such as cancer, lymphoedema and various inflammatory conditions. Recent biological and technological developments in lymphatic vascular biology will lead to a better understanding and treatment of these diseases.

Oxygen, nutrients and hormones are delivered to tissues by blood vessels, and capillaries are involved in the molecular exchange of these compounds with the surrounding tissues. Blood pressure causes plasma to leak continuously from the capillaries into the interstitial space. The main function of the lymphatic vasculature is to return this protein-rich fluid back to the circulating blood. Fluid, macromolecules and cells, such as extravasated leukocytes and activated antigen-presenting cells, enter the blind-ended lymphatic capillaries. From here, lymph is transported towards collecting lymphatic vessels and is returned to the blood circulation through the lymphatico-venous junctions in the jugular area (Fig. 1a, b). On its way, lymph is filtered through the lymph nodes, where foreign particles taken up by antigen-presenting cells are used to initiate specific immune responses. In the small intestine, lacteal lymphatic vessels inside the intestinal villi

absorb the dietary fat released by enterocytes in the form of lipid particles called chylomicrons. Lymphatic capillaries are present in the skin and in most internal organs, with the exception of the central nervous system, bone marrow and avascular tissues such as cartilage, cornea and epidermis. The lymphatic vascular system is a characteristic feature of higher vertebrates, whose complex cardiovascular system and large body size require the presence of a secondary vascular system for the maintenance of fluid balance (Box 1).

The lymphatic capillaries are thin-walled, relatively large vessels, composed of a single layer of endothelial cells. Lymphatic capillaries are not ensheathed by pericytes or smooth muscle cells, and have little or no basement membrane (Fig. 1c). Collecting lymphatic vessels have a smooth muscle cell layer, basement membrane and valves. The contraction of smooth muscle cells, and surrounding skeletal muscles, as

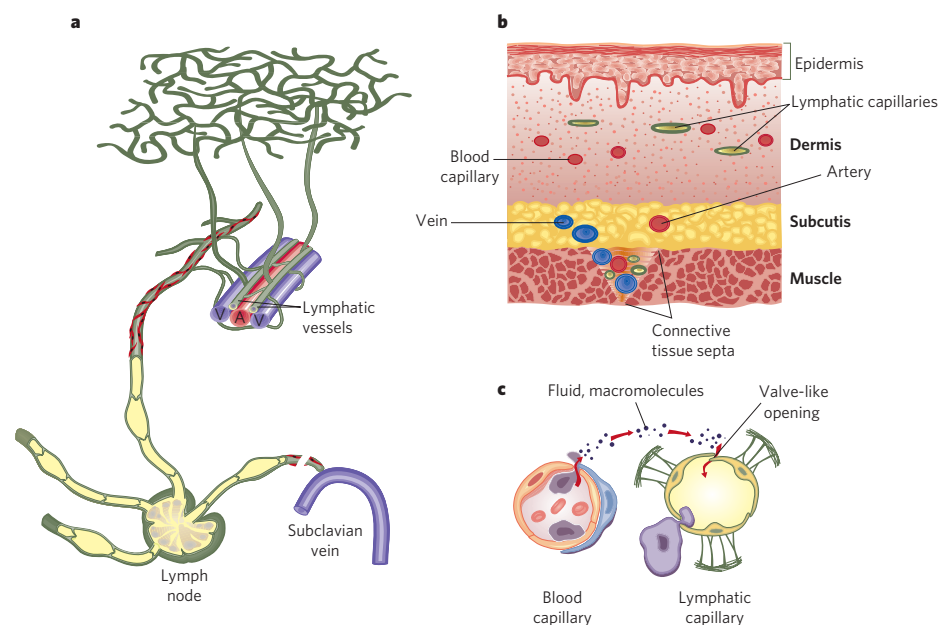


Figure 1 | Organization of lymphatic vasculature. **a**, Interstitial fluid, collected by the initial lymphatic capillary plexus, is transported by pre-collector lymphatic vessels to larger collecting lymphatic vessels and returned to the circulation through the thoracic duct. Collecting lymphatic vessels have smooth muscle cell coverage (red) and luminal valves to propel and maintain unidirectional lymph flow. Deep lymphatic vessels run along arteries and veins. **b**, Schematic cross-section of skin, showing the relative positions of blood and lymphatic vessels. **c**, Mechanism of interstitial tissue fluid uptake by a lymphatic capillary. Plasma components, extravasated white blood cells and particulate matter, such as bacteria, enter the lymphatic vessels through loose valve-like openings. Lymphatic vessels are linked to the extracellular matrix by anchoring filaments. The latter are very thin (4–10 nm) fibrillin-containing filaments, which are inserted into the endothelial cell plasma membrane. Anchoring filaments prevent vessel collapse in conditions of high interstitial pressure.

¹Molecular/Cancer Biology Laboratory, Ludwig Institute for Cancer, Research and ²Development and Differentiation Laboratory, Molecular/Cancer Biology Program, Biomedicum Helsinki, Haartman Institute and Helsinki University Central Hospital, P.O.B. 63 (Haartmaninkatu 8), 00014 University of Helsinki, Finland.

Box 1 | Evolutionary aspects of lymphatic vascular development

Molluscs and arthropods have an open circulatory system, which combines both blood and lymphatic functions. Blood (or haemolymph) freely diffuses to the tissues to distribute oxygen, nutrients and to collect metabolic wastes. Vertebrates and some invertebrates have closed circulatory systems, where blood is always contained within vessels. With increasing animal size and associated complexity of the cardiovascular system, an additional mechanism is necessary to clear the tissue of the proteins, microbes and other substances that blood vessels are unable to absorb. Lymphatic vessels are present in amphibia and reptiles. An interesting feature of the lymphatic system in several of these animals is the presence of enlarged, rhythmically contracting portions of lymphatic vessels called lymph hearts. Lymph hearts contain valves and are necessary to propel lymph and to prevent lymph backflow. Recently, the *Xenopus* tadpole has become the system of choice for the developmental genetic analysis of lymphangiogenesis. For example, prox-1-GFP transgenic tadpoles can be used to visualize lymphatic vessels *in vivo*, and morpholino oligonucleotides can be used to downregulate lymphangiogenic genes (P. Carmeliet, personal communication; ref. 96). In birds and amphibians, Prox1-expressing lymphangioblasts also develop independently in the mesenchyme^{96,97}. However, it is not known whether similar precursors exist in mammals. At least in mouse xenograft experiments, very few, if any, bone-marrow-derived cells are incorporated into growing lymphatic vessels⁹⁸. Lymph nodes are present only in mammals and some aquatic birds; their number is significantly lower in birds.

well as arterial pulsations, contribute to lymph propulsion, and valves prevent backflow.

Lymphatic vessels were first described in the beginning of the seventeenth century; however, the first growth factors and molecular markers specific for these vessels were discovered only ten years ago. In retrospect, this may seem surprising, given the well-known importance of the lymphatic system in maintaining the fluid balance in the body, and its involvement in the pathogenesis of many diseases, including cancer. Recent developments in lymphatic vascular biology research include the discovery of lymphangiogenic factors, identification of lymphatic vascular markers, isolation of lymphatic endothelial cells and the development of animal models to study lymphangiogenesis. The molecular mechanisms of lymphatic growth and development have been recently reviewed¹⁻³. In this review, we summarize the recent progress in this fast-growing field of vascular biology with particular emphasis on the understanding and management of lymphatic dysfunction, inflammation and tumour metastasis.

Mechanisms of lymphangiogenesis

Studies over the past ten years have revealed a signal-transduction system for lymphatic endothelial cell growth, migration and survival. This system is formed by vascular endothelial growth factors (VEGF) C and D and their receptor VEGFR-3 (Fig. 2a, b)⁴⁻⁸. VEGF-C and VEGF-D also bind to neuropilin-2 (Nrp2), a semaphorin receptor in the nervous system that is also expressed in lymphatic capillaries⁹. Consistent with these findings, Nrp2-deficient mice have lymphatic hypoplasia¹⁰. Proteolytically processed VEGF-C and VEGF-D also activate VEGFR-2 and can induce blood-vessel growth^{5,11-14}. Conversely, VEGF-A, which binds to VEGFR-2, can induce lymphatic hyperplasia but cannot substitute for VEGF-C in lymphatic development^{15,16}. By contrast, in the chick chorioallantoic membrane and in a mouse insulinoma tumour model, VEGF-A stimulates only angiogenesis^{17,18}. At least some of the effects of VEGF-A on lymphatic vessels may be secondary to the induction of vascular hyperpermeability and to the recruitment of the inflammatory cells that produce VEGF-C and VEGF-D^{19,20}.

The recent identification of co-receptors and novel signalling complexes for lymphangiogenic signalling suggests a greater complexity than previously thought. *In vitro* studies show that upon binding to matrix fibronectin, β 1 integrin interacts with VEGFR-3 and induces weak activation of its tyrosine kinase²¹. Integrin α 9 binds VEGF-C and inactivation of *Itga9* causes chylothorax in mice^{22,23}. Furthermore,

Kaposi sarcoma herpesvirus envelope glycoprotein gB interacts with VEGFR-3 and α 3 β 1 integrin and can activate both, resulting in increased endothelial cell growth and migration²⁴ (Box 2; Fig. 2b). In addition to the two VEGF family members, fibroblast growth factor 2, platelet-derived growth factor B and hepatocyte growth factor stimulate lymphatic vessel growth²⁵⁻²⁷.

Mechanisms of embryonic and postnatal lymphangiogenesis

In humans, lymph sacs appear in 6–7-week-old embryos, and in the mouse, lymph-vessel development begins around embryonic day 10 (E10). So far, experimental data from mice support the hypothesis proposed by Florence Sabin about 100 years ago that lymphatic endothelial cells arise by sprouting from embryonic veins in the jugular and perimesonephric areas. From here they migrate to form primary lymph sacs and the primary lymphatic plexus, which is composed of capillary-like vessels²⁸ (Fig. 3; Table 1). The homeobox transcription factor Prox1 and VEGF-C are essential for these initial developmental events. These and other factors involved are addressed below.

Prox1

In mice, Prox1-expressing endothelial cells are first observed at E10 in the jugular vein, from which they migrate to form the first lymphatic sprouts²⁹. *Prox1* deletion leads to a complete absence of the lymphatic vasculature; endothelial cells bud from the cardinal vein but fail to express lymphatic endothelial markers and do not migrate further (Fig. 3)³⁰. Accordingly, PROX1 overexpression in human blood vascular endothelial cells suppresses many blood vascular-specific genes and upregulates lymphatic endothelial-cell-specific transcripts^{31,32}. At present, the signals leading to the polarized expression of Prox1 and its target genes in lymphatic endothelial cells are not known. *Prox1*^{+/-} mice die perinatally in most genetic backgrounds except in the outbred NMRI background, in which they develop chylous ascites and adult-onset obesity³³. Notably, endothelium-specific deletion of *Prox1* at least partly recapitulates the obese phenotype, indicating a link between abnormal lymphatic vessel development, impaired lymph drainage and tissue adiposity³³.

VEGF-C/D and VEGFR-3

Homozygous deletion of *Vegfc* leads to the complete absence of the lymphatic vasculature in mouse embryos, whereas *Vegfc*^{+/-} mice display severe lymphatic hypoplasia¹⁶. In *Vegfc*-null mice, lymphatic endothelial cells initially differentiate in the cardinal veins but fail to migrate and to form primary lymph sacs. This demonstrates that VEGF-C is an essential chemotactic and survival factor during embryonic lymphangiogenesis¹⁶. By contrast, deletion of *Vegfd* does not affect development of the lymphatic vasculature, although exogenous VEGF-D protein rescues the impaired vessel sprouting in *Vegfc*^{-/-} embryos^{16,34}. *Vegfr3* deletion leads to defects in blood-vessel remodeling and embryonic death at mid-gestation, indicating an early blood vascular function³⁵.

Box 2 | Origin of Kaposi sarcoma spindle cells

Kaposi sarcoma is a neoplasm characterized by vascular nodules in the skin, mucous membranes and internal organs. It is endemic in sub-Saharan regions in Africa and is frequently encountered in AIDS patients. The nodules are composed of clusters of spindle-shaped tumour cells and characterized by a prominent vasculature. The spindle cells express both blood and lymphatic endothelial cell markers, suggesting their endothelial origin. Development of Kaposi sarcoma is associated with infection by the human herpesvirus-8 (HHV-8). The transcriptional profile of Kaposi sarcoma cells is closely related to normal lymphatic endothelial cells^{99,100}. Furthermore, *in vitro* infection of blood vascular endothelial cells with HHV-8 resulted in the expression of several lymphatic endothelial cell-specific genes, although HHV-8-infected lymphatic endothelial cells also showed some infidelity of phenotypic gene expression^{99,100}. Growth signals may also be provided directly by HHV-8 through interactions between one of its capsid proteins and VEGFRs or integrins on endothelial cells²⁴.

Heterozygous missense point mutations, which lead to tyrosine-kinase inactivation, have been found in *VEGFR3* in patients with Milroy disease (OMIM 153100), a rare autosomal dominant lymphoedema characterized by hypoplasia of cutaneous lymphatic vessels³⁶. Chy mice, derived from an ethylnitrosourea mutagenesis screen, have a similar mutation and develop lymphoedema. They are a useful model for studies of hereditary lymphoedema and its therapy⁹.

LYVE-1

LYVE-1 (lymphatic vessel hyaluronan receptor-1) is one of the most widely used markers for lymphatic endothelial cells³⁷. LYVE-1 is the first marker of lymphatic endothelial commitment, and in mice it is expressed in a polarized manner in venous endothelium starting from E9 (Fig. 3). In adults, LYVE-1 expression is downregulated in the collecting lymphatic vessels but remains high in lymphatic capillaries³⁸. The role of LYVE-1 in the regulation of lymphatic vascular function is not known, but mice lacking this receptor have normal lymphatic vessels (G. Thurston, personal communication).

Syk and SLP76

A connection to the thoracic duct at the junction of the left subclavian and the internal jugular veins is the main point of entry of lymph to the bloodstream. Additional lymphatico-venous communications occur in the renal, hepatic and adrenal veins, in the lymph nodes and in other

peripheral locations^{39,40}. Lymphatico-venous anastomoses are frequently observed in lymphoedema, chylous ascites and chylothorax, where they are an adaptive response to lymphatic hypertension.

The tyrosine kinase Syk and adaptor protein SLP76 are involved in controlling the separation of the lymphatic and blood vascular systems. Mice with mutations in these molecules have arterio-venous shunts and abnormal lymphatico-venous communications. Syk and SLP76 are expressed almost exclusively in haematopoietic cells, suggesting that these cells contribute to the separation of the two vascular systems⁴¹.

Further development of the lymphatic vessels involves remodelling during late embryogenesis and postnatally to form a superficial capillary plexus and collecting lymphatic vessels. Genetic ablation experiments point to the involvement of several genes in this process; these are highlighted below (also see Table 1).

Angiopoietins and Tie receptors

Little is known of the functions of the angiopoietins (Ang) in the lymphatic vasculature. The Ang receptors Tie1 and Tie2 are expressed by lymphatic endothelial cells⁴², and Ang2 is considered to be an endogenous inhibitor of Tie2 in endothelial cells, although in some conditions it can act like the agonistic Ang1. *Angpt2*-gene-targeted mice display lymphatic hypoplasia, but replacement of *Angpt2* with *Angpt1* was sufficient to rescue the lymphatic vascular phenotype⁴³. Furthermore,

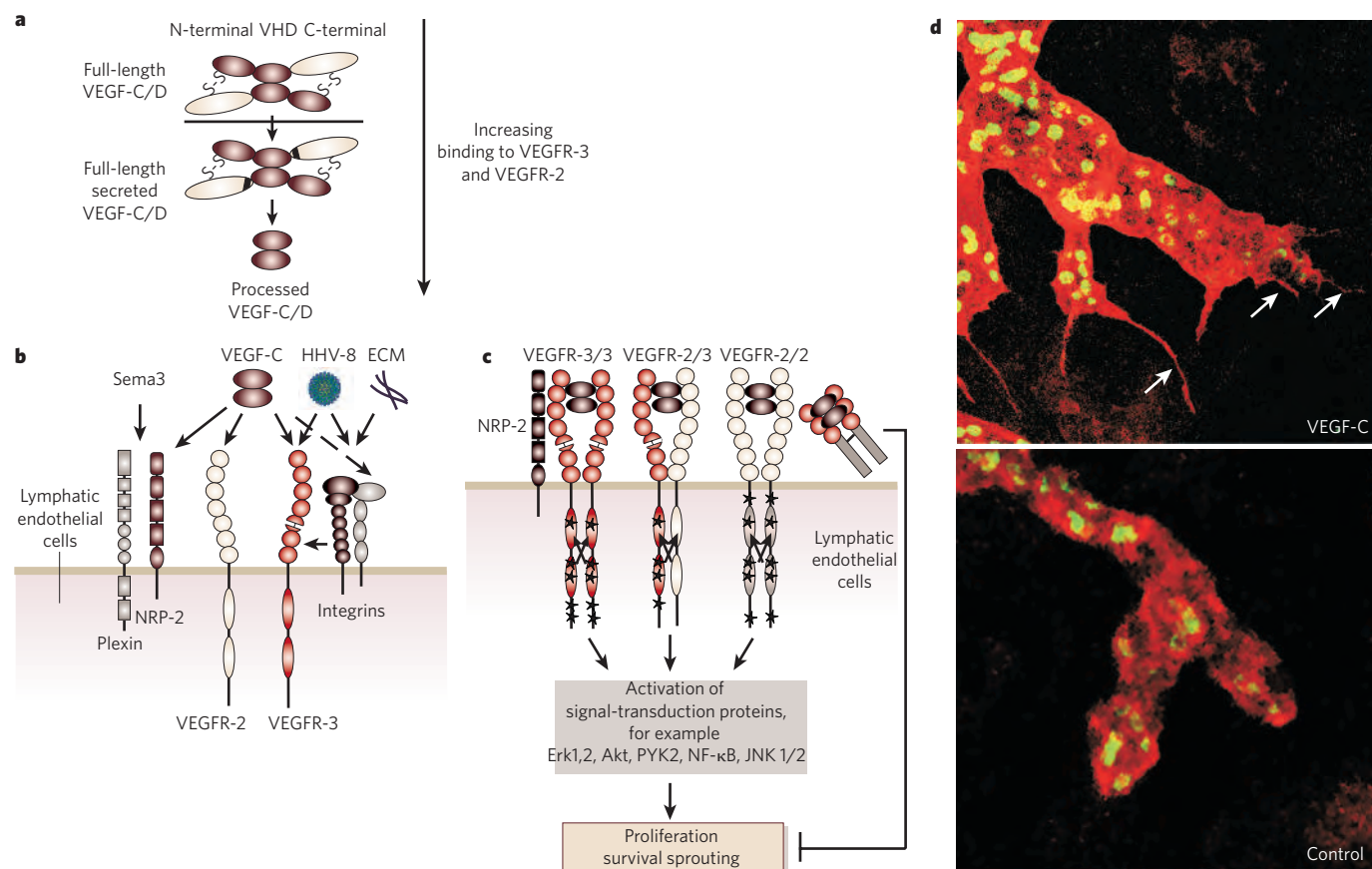


Figure 2 | VEGF-C/D-VEGFR-3 pathway in the regulation of the lymphatic vessel growth. **a**, Stepwise proteolytic processing of the VEGF-C or VEGF-D homodimers results in gradually increased binding to VEGFR-3 and VEGFR-2. VEGF-C and VEGF-D are activated by intracellular secretory proprotein convertases. The secreted subunits of these factors are disulphide bonded by means of their propeptides, but they are further proteolysed in the extracellular environment to generate non-disulphide-linked homodimeric proteins (reviewed in ref. 39). VHD, VEGF homology domain. **b**, Extracellular matrix proteins (ECM), such as collagen and fibronectin, enhance tyrosine phosphorylation of VEGFR-3 through activation of integrin $\beta 1$. Additional, less explored VEGF-C signal-transduction pathways include interaction of VEGF-C with NRP-2 and integrin $\alpha 9$, and VEGFR-3 with integrin $\beta 1$ and HHV-8 envelope protein $gB^{9,21,22,24}$. NRP-2 also serves as a plexin co-receptor for type III semaphorins (Sema 3). The importance of these interactions for the *in vivo* signalling through VEGFR-3 still needs to be demonstrated. **c**, Formation of homo- and heterodimeric VEGFR-2 and VEGFR-3 receptor complexes leads to tyrosine phosphorylation (stars), the recruitment of intracellular signal-transduction proteins, and enhanced endothelial cell migration, proliferation and survival^{94,95}. **d**, Adenoviral delivery of VEGF-C induces lymphangiogenesis in mouse skin. Lymphatic vessels are visualized by staining for VEGFR-3 (red) and PROX1 (green). Arrow points at sprouts of growing lymphatic vessels.

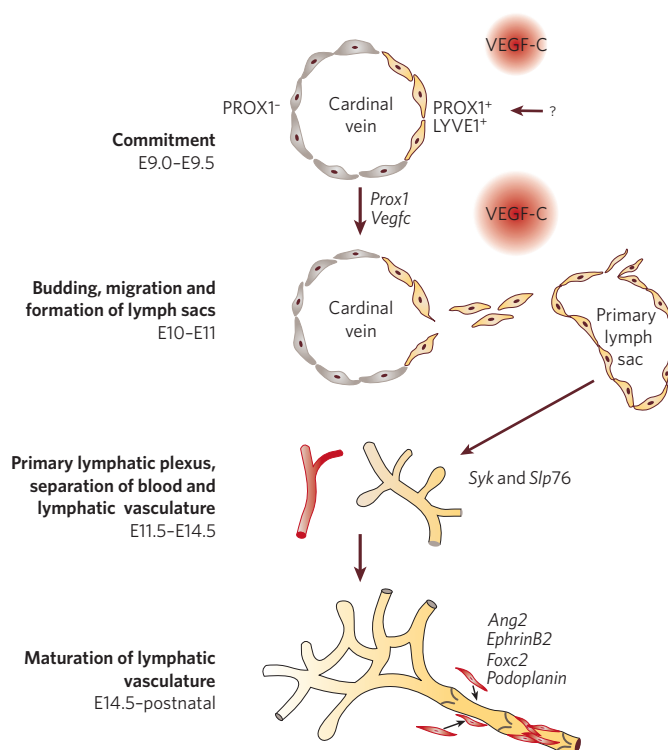


Figure 3 | Model for the development of mouse lymphatic vasculature (see explanations in the main text). For more detailed discussion of developmental lymphangiogenesis see ref. 2.

Ang1 induces lymphatic vessel growth in adult tissues^{44,45}. It is unclear how the angiopoietins convey lymphangiogenic signals.

EphrinB2

Postnatal remodelling of the lymphatic vasculature includes sprouting of lymphatic capillaries from the primary lymphatic plexus, whereas deeper lymphatic vessels recruit smooth muscle cells and develop lymphatic valves, acquiring a collecting vessel phenotype³⁸. Mice expressing a mutated form of the transmembrane growth factor ephrinB2, which lacks the carboxy-terminal site for binding PDZ-domain-containing proteins, have a normal blood vasculature but display a disturbed postnatal remodelling of the lymphatic vasculature. This leads to hyperplasia of the collecting lymphatic vessels, lack of luminal valve formation and a failure to remodel the primary lymphatic capillary plexus³⁸. The ephrins and their Eph receptors have been implicated in repulsive axon guidance in the nervous system and in the control of blood-vessel remodelling (ref. 46; see also the review by Coultas, Chawengsaksophak and Rossant in this issue, p. 937). The new data suggest that there are interesting differences in the remodelling processes between blood and lymphatic vascular systems.

Foxc2

The forkhead transcription factor Foxc2 is involved in the specification of the lymphatic capillary versus collecting lymphatic vessel phenotype. Foxc2 is highly expressed in the developing lymphatic vessels as well as in lymphatic valves in adults^{47,48}. The early development of lymphatic vessels proceeds normally in the absence of Foxc2, but later the patterning of lymphatic vasculature becomes abnormal. Moreover, collecting lymphatic vessels in *Foxc2*^{-/-} mice lack valves, whereas the lymphatic capillaries acquire ectopic coverage by basal lamina components and smooth muscle cells⁴⁷.

Ectopic smooth muscle cells surrounding abnormal lymph vessels are also found in humans suffering from lymphoedema-distichiasis (LD, OMIM 153400), an autosomal dominant disease caused by heterozygous loss-of-function mutations of *FOXC2* (ref. 49). LD is char-

acterized by late-onset lymphoedema and a double row of eyelashes. Unlike in Milroy disease, the lymphatic vasculature in LD is normal or hyperplastic, but there is lymph backflow, presumably due to abnormal lymphatic valves, defective vessel patterning and the presence of ectopic smooth muscle cells^{47,50}. Many LD patients also suffer from incompetent venous valves (P. Mortimer, personal communication), suggesting that *FOXC2* is also important for their development.

Podoplanin

Podoplanin is transmembrane mucin-type glycoprotein that is highly expressed in podocytes, keratinocytes, cells of choroid plexus, alveolar lung cells and lymphatic endothelial cells. Podoplanin deficiency leads to abnormal lung development and perinatal lethality. Podoplanin knockout mice displayed paw lymphedema and abnormal lymphatic function and patterning, perhaps due to impaired migration of lymphatic endothelial cells⁵¹.

Molecular blueprint of lymphatic endothelial cells

The discovery of cell-surface markers, such as VEGFR-3, podoplanin, LYVE-1 and CD34, that distinguish blood vascular from lymphatic endothelial cells has allowed the isolation of pure populations of these two cell types from human skin⁵²⁻⁵⁵. Growth of cultured lymphatic endothelial cells is dependent on VEGF-C, which in mixed cultures is supplied by the blood vascular endothelial cells. Interestingly, both cell types show preferentially homotypic interactions, even *in vitro*⁵². Approximately 2% of transcribed genes are differentially expressed between lymphatic and blood vascular endothelial cells, and this difference may reflect their distinct *in vivo* functions^{32,54,55}. Detailed discussion of the expression-profiling studies has been provided in recent reviews^{1,3}. Although the transcripts expressed by lymphatic and blood vascular endothelial cells are significantly modified soon after their isolation from tissues (P. Saharinen and N. Wick, personal communication), several genes potentially important in the regulation of lymphatic vascular function have been identified. Further analysis of their functions by gene knockout and knockdown should provide a comprehensive view of lymphatic vascular biology in the coming few years.

Lymphatic vascular insufficiency and its treatment

Impairment of the lymphatic-transport capacity because of abnormal vessel development or damaged lymphatic vessels causes stagnation of proteins and associated water in the interstitium, and leads to lymphoedema, usually a progressive and lifelong condition for which no curative treatment exists. The protein-rich interstitial fluid initiates an inflammatory reaction, leading to fibrosis, impaired immune responses and fatty degeneration of the connective tissue. Although primary, congenital lymphoedema is commonly the result of inherited genetic damage, secondary lymphoedema is caused by filariasis (elephantiasis) or by traumas due to radiation therapy, surgery or infection. Filariasis is a parasitic infection of lymphatic vessels by *Wuchereria bancrofti* or *Brugia malayi* worms, transmitted by mosquito bites. This leads to damage of lymphatic vessels and chronic lymphoedema of legs or genitals. Filariasis is the main cause of lymphoedema in tropical countries, with some 100 million people affected worldwide, whereas breast-cancer surgery is a leading cause for secondary lymphoedema in industrialized countries⁵⁶.

Chylous ascites and chylothorax are caused by accumulation of high-fat-containing fluid or chyle in the abdomen or thorax as a result of trauma, obstruction or abnormal development of lymphatic vessels⁴⁰. This leads to lymphatic hypertension, lymph extravasation and loss of proteins, lipids and leukocytes as well as abdominal inflammation and intestinal adhesions. Chylous ascites or chylothorax may accompany other types of lymphatic vascular dysfunction, especially in mouse models (see Table 1), whereas peripheral lymphoedema is often inconspicuous in these animals, probably due to their small size and the low hydrostatic pressure in the limbs.

Recently, promising lymphoedema treatment results have been

Table 1 | Genes that mediate lymphatic vasculature formation and patterning

Gene	Function	Lymphatic phenotype	Other defects	Lethality
<i>Angpt2</i> , gene targeted ⁴³	Growth factor, ligand of Tie-2	Hypoplasia, chylous ascites ($^{-/-}$)	Eye hyaloid vasculature fails to regress ⁴³ , abnormal inflammatory response (H. Augustin, personal communication)	$^{-/-}$: perinatal or normal
<i>Efnb2</i> , PDZ-binding mutant ³⁸	Ligand of EphB receptors	Retrograde lymph flow, chylothorax, ectopic mural cells, absent valves ($\Delta V/\Delta V$)	Lung development defects (R. Klein, personal communication)	$^{-/-}$: perinatal $+/+$: normal
<i>Foxc2</i> , gene targeted	Transcription factor	Abnormal lymphatic patterning, presence of mural cells, absent valves ($^{-/-}$) ⁴⁷ , lymphatic vessel and lymph node hyperplasia ($+/+$) ⁸⁹	Aortic arch malformations, heart septal defects, abnormal kidney and urethra development	$^{-/-}$: E12.5-perinatal $+/+$: normal
<i>Itga9</i> , gene targeted ²³	Adhesion receptor	Lymphoedema, chylothorax ($^{-/-}$)	Not reported	$^{-/-}$: perinatal $+/+$: normal
<i>Elk3</i> (Net), gene targeted ⁹⁰	Transcription factor Mutant form lacks DNA-binding domain	Lymphangiectasis, chylothorax ($^{-/-}$)	Impaired wound and tumour angiogenesis $^{+/-}$	$^{-/-}$: perinatal $+/+$: normal
<i>Nrp2</i> , gene targeted ¹⁰	Receptor for VEGF165, VEGF145, PIGF, VEGF-C and class 3 semaphorins	Transient hypoplasia of lymphatic capillaries ($^{-/-}$)	Defects in neural fasciculation and guidance	$^{-/-}$: perinatal $+/+$: normal
Podoplanin (<i>Gp38</i>), gene targeted ⁵¹	Membrane glycoprotein	Lymphangiectasis, abnormal lymph transport, lymphoedema ($^{-/-}$)	Respiratory failure due to abnormal lung development	$^{-/-}$: perinatal $+/+$: normal
<i>Pik3r1</i> , gene targeted ⁹¹	Regulatory subunits of class I _A PI(3)K	Chylous ascites ($^{-/-}$)	Liver necrosis, enlarged skeletal muscle fibres, brown fat depositions, calcification of heart tissue	$^{-/-}$: perinatal (129Sv $\times$ C57Bl6) or 30% survival (outbred) $+/+$: normal
<i>Prox1</i> , gene targeted or endothelial specific deletion ^{29,33}	Transcription factor	No lymphatic vessels ($^{-/-}$) Chylous ascites, adult onset obesity ($+/+$)	Abnormal eye, liver and pancreas development	$+/+$: perinatal in most backgrounds $^{-/-}$: E14.5
<i>Lcp2</i> (SLP-76) and <i>Syk</i> , gene targeted ⁴¹	Tyrosine kinase (Syk); adaptor protein (SLP-76)	Failure of separation of blood and lymphatic vasculature, chylous ascites ($^{-/-}$)	Failure of T-cell development and fetal haemorrhage (Slp76) Block of B-cell development and fetal haemorrhage (Syk)	<i>Lcp2</i> $^{-/-}$: perinatal <i>Lcp2</i> $^{+/-}$: normal <i>Syk</i> $^{-/-}$: perinatal <i>Syk</i> $^{+/-}$: normal
<i>Sox18</i> (ragged) ⁹²	Transcription factor Spontaneous missense mutations	Edema and chylous ascites ($^{-/-}$)	Lack of vibrissae and coat hairs, generalized oedema and cyanosis due to cardiovascular defects	$^{-/-}$: perinatal $+/+$: normal
Trisomy 16 (Ts16) ⁹³	Many	Nuchal oedema, abnormal size and structure of jugular lymph sacs from E14	Multiple cardiac or craniofacial development defects	E16-E20
<i>Vegfc</i> , gene targeted ¹⁶	Growth factor, ligand of VEGFR-3	No lymphatic vessels ($^{-/-}$) hypoplasia, chylous ascites ($+/+$)	Not reported	$^{-/-}$: E17-E19 $+/+$: perinatal or normal
<i>Vegfr3</i> (<i>Chy</i> , ethylnitrosourea-induced mutation) ⁹	Receptor tyrosine kinase, kinase-inactivating mutation I1053F	Hypoplasia, chylous ascites ($+/+$)	Failure of remodelling of primitive blood vascular plexus ($^{-/-}$)	$^{-/-}$: E10 $+/+$: perinatal or normal

achieved in preclinical models using viral gene-transfer vectors that encode lymphangiogenic growth factors (reviewed in ref. 57). For example, *VEGF-C* gene-transduction induces growth of functional lymphatic vessels⁵⁸, whereas the mature form of *VEGF-D* is a very powerful inducer of angiogenesis¹⁴. Lymphatic vascular growth without concomitant blood vascular side effects was selectively induced with the VEGFR-3-specific ligand VEGF-C156S (ref. 58). VEGF-C gene therapy was effective even in *Chy* mice that suffer from lymphoedema caused by a heterozygous inactivating mutation of VEGFR-3 (ref. 9). *ANG1* gene transfer to mouse skin promoted lymphangiogenesis, simultaneously inhibiting vascular hyperpermeability. This factor could also be used for the treatment of tissue oedema^{45,59}. The pathophysiology of vascular permeability has been recently reviewed elsewhere⁶⁰.

Tumour metastasis to lymph nodes and its inhibition

Metastatic tumour spread through the blood or lymphatic vessels occurs in most forms of human cancer, with regional lymph-node metastasis often being the most important prognostic factor for carcinoma patients⁶¹. From the sentinel lymph node, which is the first regional lymph node to which tumour cells metastasize, further dissemination may occur to other nodes and distant organs. At present, it is not clear whether lymphatic metastasis selects cells with increased

potential for subsequent organ metastasis or simply indicates that the tumour has become metastatic in general.

Growth-factor stimulation of lymphatic vessels enhances lymphatic metastasis. Several studies have found positive correlations between VEGF-C or VEGF-D expression and vascular invasion, lymphatic vessel and lymph node involvement, distant metastasis and, in some instances, poor clinical outcomes⁶². VEGF-C expression in tumour cells may be induced by growth factors or proinflammatory cytokines, and some may be derived from inflammatory cells in tumours. High levels of VEGF-C or VEGF-D also enhance lymphatic metastasis in various experimental models⁶³⁻⁶⁷. Furthermore, in some tumours, proteolytically processed VEGF-C or VEGF-D may be generated, which targets VEGFR-2 or VEGFR-3 that is often upregulated in tumour blood vessels⁶⁸. A direct link between VEGF-C or VEGF-D expression and metastasis was established with the use of a soluble VEGFR-3-immunoglobulin fusion protein (VEGF-C/D trap) or blocking anti-VEGF-D antibodies^{63,66,67}. In some models lymphatic, but not lung, metastases were blocked with the VEGF-C/D trap, whereas in others the treatment inhibited both types of metastases^{63,69}. Although these experiments provide support for the contribution of VEGF-C, VEGF-D, and their receptor, VEGFR-3, in lymphatic spread in malignancy, the mechanisms of these effects have only recently been addressed.

Proliferating intratumoural lymphatic vessels are present in certain

human cancers, such as melanomas, head and neck carcinomas and xenograft tumour models overexpressing lymphangiogenic factors^{70,71}. However, they may not be a prominent feature, and may in fact not be required for enhanced metastasis in most solid tumours. At least in animal models, intratumoural lymphatic vessels may not be completely functional, because these vessels collapse under high intratumoural pressure⁷², and at least in one study they were not conducive of lymphatic metastasis⁷³. We favour the view that local lymphatic vessels at the tumour margin are more important for spreading tumour cells, through the process of vessel sprouting under the influence of interstitial fluid hypertension and tumour-secreted VEGF-C^{74,75}. In this process endothelial cells send long filopodia towards the VEGF-C-producing tumour cells and then form tumour-directed vessel sprouts, where the vessel lumen opens up and allows facilitated access of tumour cells to the lumen (Fig. 4). Tumour lymphatic vessels carry specific markers, such as CD34, and their heterogeneity can be used for their targeting (H. Augustin, personal communication; ref. 76).

Some evidence indicates that the lymphatic endothelium actively participates in metastasis formation by secreting chemokines, such as CCL21 (SLC, 6CKine and Exodus), whose receptor (CCR7) is expressed on some tumour cells⁷⁷. Furthermore, the collecting lymphatic vessels draining fluid from the tumour area are stimulated by intraluminal VEGF-C to dilate through the process of endothelial proliferation in the vessel wall⁷⁵. Clumps of metastatic tumour cells could then undergo an easier transit in lymph, flowing in the dilated hyperplastic vessels. The VEGF-C/D trap inhibited the sprouting and vessel dilation and seemed to restore the integrity of the vessel wall⁷⁵. Similarly, blocking monoclonal antibodies that target VEGF-C, VEGF-D or their receptor(s) and small molecules that inhibit the tyrosine kinase catalytic domain of these receptors could be used for the inhibition of tumor metastasis. Further work should soon tell if these same molecules inhibit further systemic metastasis or angiogenesis in some tumours. In this case such compounds would undoubtedly proceed to clinical trials. However, it should be noted that also VEGF can stimulate lymphatic metastasis⁷⁸.

phatic vessels draining fluid from the tumour area are stimulated by intraluminal VEGF-C to dilate through the process of endothelial proliferation in the vessel wall⁷⁵. Clumps of metastatic tumour cells could then undergo an easier transit in lymph, flowing in the dilated hyperplastic vessels. The VEGF-C/D trap inhibited the sprouting and vessel dilation and seemed to restore the integrity of the vessel wall⁷⁵. Similarly, blocking monoclonal antibodies that target VEGF-C, VEGF-D or their receptor(s) and small molecules that inhibit the tyrosine kinase catalytic domain of these receptors could be used for the inhibition of tumor metastasis. Further work should soon tell if these same molecules inhibit further systemic metastasis or angiogenesis in some tumours. In this case such compounds would undoubtedly proceed to clinical trials. However, it should be noted that also VEGF can stimulate lymphatic metastasis⁷⁸.

Lymphangiogenesis in inflammation

Lymphatic vessels proliferate during inflammation⁷⁹. Pro-inflammatory cytokines induce VEGF-C messenger RNA transcription, presumably through NF- κ B-mediated promoter activation, suggesting that they regulate lymphatic vessel growth during inflammation⁸⁰. Interestingly, constitutive NF- κ B activity is detected in the lymphatic endothelium *in vivo*, but its role in the lymphatic endothelial cells remains enigmatic⁸¹.

Inflammatory infiltrates in human kidney transplants undergoing rejection contain proliferating host lymphatics⁸². Infection of mouse airway epithelial cells with the respiratory pathogen *Mycoplasma pulmonis* resulted in robust lymphangiogenesis driven by VEGF-C- and VEGF-D-expressing immune cells that could be inhibited by using a VEGF-C/D trap²⁰. Importantly, VEGFR-3 inhibition resulted in severe exacerbation of mucosal oedema and reactive lymphadenitis decreased. This is consistent with the importance of the lymphatic vascular system as an exit route for immune cells and fluid²⁰. In a rabbit cornea model of inflammatory angiogenesis and lymphangiogenesis, either a VEGF inhibitor or selective depletion of the VEGF-C and VEGF-D producing macrophages blocked lymphangiogenesis, demonstrating that inflammatory cells recruited by VEGF can mediate the formation of lymphatic vessels¹⁹. Moreover, dendritic cells expressing both VEGFR-3 and VEGF-C could be detected in a mouse model of corneal inflammation, suggesting that immune cells may both respond to lymphangiogenic signals and induce lymphangiogenesis (ref. 83). Indeed, blockade of VEGFR-3 signalling suppressed trafficking of corneal dendritic cells to draining lymph nodes and inhibited induction of delayed-type hypersensitivity and rejection of corneal transplants⁸⁴.

Lymphatic vessels participate in the regulation of inflammatory response through their role in transport of lymphocytes to the lymph nodes. Migration of dendritic cells is mediated by the chemokine receptor CCR7, whereas lymphatic vessels express the ligand CCL21⁸⁵. Furthermore, mannose receptor 1 and common lymphatic endothelial and vascular endothelial receptor-1 (CLEVER-1) control lymphocyte traffic in lymphatic vessels^{86,87}. Human lymphatic endothelial cells also express the D6 chemokine receptor, which is involved in the post-inflammatory clearance of beta-chemokines⁸⁸.

Future directions

Recent progress in the area of lymphatic vascular biology has provided various genetic mouse models and new molecular tools for isolation and growth regulation of lymphatic vessels. Coupled with high-throughput genomic, proteomic and functional screens, these methods will undoubtedly reveal additional possibilities for therapeutic intervention in diseases where the lymphatic vascular system has a significant pathophysiological function. Below we have briefly outlined the questions that we believe should be addressed in the next few years.

The early steps of lymphatic endothelial cell commitment are not yet understood, and the mechanisms of lymphatic vascular remodeling, patterning and maturation are only beginning to be elucidated. Studies of blood vascular development have shown that Notch,

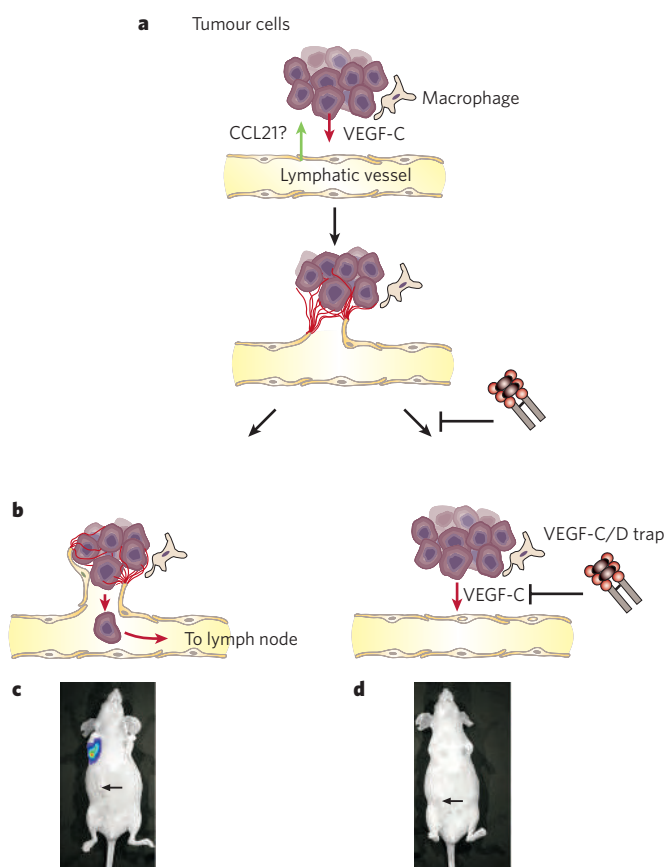


Figure 4 | Role of VEGF-C/D in lymphatic metastasis in cancer. **a**, Tumour cells and tumour-associated macrophages secrete lymphangiogenic growth factor VEGF-C or VEGF-D, which induces sprouting of nearby lymphatic vessels, facilitating the access of tumour cells into the vessel lumen. The lymphatic endothelial cells may also actively attract some tumour cells through the secretion of chemokines, such as CCL21. **b**, Aggregates of tumour cells are transported to the regional lymph node, from which they can spread to distant organs through either blood or lymphatic vessels. Blockage of VEGFR-3 signalling inhibits metastasis in most mouse tumour xenograft models by stabilizing lymphatic vessels. **c**, Nude mice were implanted with luciferase-tagged tumour cells, which metastasize to the ipsilateral axillary lymph node in control adenovirus-treated mice. **d**, By contrast, metastasis was abolished in mice treated with adenovirus encoding a VEGF-C/D trap⁷⁵. Note that the primary tumours have been excised (arrows).

Eph/ephrin, Shh and TGF- β pathways have an important role in the specification of arterial versus venous cell fates, whereas neural-guidance molecules such as netrins, semaphorins, plexins and members of Slit/Robo family are essential for vessel remodelling and navigation (see also p. 937). Furthermore, interaction of endothelial cells and pericytes, mediated in part through PDGF-B/PDGFR β , is necessary during blood-vessel maturation. It will be important to determine which signalling pathways control different stages of lymphatic vascular development and to what extent they are similar to the ones operating in the blood vessels.

Lymphangiogenesis research has so far provided imminent therapeutic applications for human diseases such as lymphoedema and other tissue oedemas that will enter clinical development in the near future. A crucial question concerns the possibility of inhibiting lymphatic metastasis in cancer patients. The importance of lymph-node metastasis in the spread of cancer to distant organs needs to be better understood before the new knowledge can be applied to patients. In this context, the possible roles of VEGF-C, VEGF-D and VEGFR-3 upregulation in tumour angiogenesis need to be explored for additional therapeutic applications.

Understanding the mechanisms of lymphatic metastasis, including the identification of stromal and tumour determinants that are important for the spread of tumour cells through lymphatic vessels, represents another challenge for tumour vascular biology researchers. Furthermore, characterization of lymphatic endothelial cells from different vascular beds including various tumour types will provide important novel targets for therapy, along with new information about normal and diseased lymphatic vascular function. Finally, the involvement of lymphatic vessels in inflammation should be explored in several contexts. ■

- Hirakawa, S. & Detmar, M. New insights into the biology and pathology of the cutaneous lymphatic system. *J. Dermatol. Sci.* **35**, 1–8 (2004).
- Oliver, G. Lymphatic vasculature development. *Nature Rev. Immunol.* **4**, 35–45 (2004).
- Tammela, T., Petrova, T. V. & Alitalo, K. Molecular lymphangiogenesis: new players. *Trends Cell Biol.* **15**, 434–441 (2005).
- Kaipainen, A. et al. Expression of the fms-like tyrosine kinase FLT4 gene becomes restricted to lymphatic endothelium during development. *Proc. Natl Acad. Sci. USA* **92**, 3566–3570 (1995).
- Achen, M. G. et al. Vascular endothelial growth factor D (VEGF-D) is a ligand for the tyrosine kinases VEGF receptor 2 (Flk1) and VEGF receptor 3 (Flt4). *Proc. Natl Acad. Sci. USA* **95**, 548–553 (1998).
- Joukov, V. et al. A novel vascular endothelial growth factor, VEGF-C, is a ligand for the Flt4 (VEGFR-3) and KDR (VEGFR-2) receptor tyrosine kinases. *EMBO J.* **15**, 290–298 (1996).
- Jeltsch, M. et al. Hyperplasia of lymphatic vessels in VEGF-C transgenic mice. *Science* **276**, 1423–1425 (1997).
- Mäkinen, T. et al. Isolated lymphatic endothelial cells transduce growth, survival and migratory signals via the VEGF-C/D receptor VEGFR-3. *EMBO J.* **20**, 4762–4773 (2001).
- Karkkainen, M. J. et al. A model for gene therapy of human hereditary lymphedema. *Proc. Natl Acad. Sci. USA* **98**, 12677–12682 (2001).
- Yuan, L. et al. Abnormal lymphatic vessel development in neuropilin 2 mutant mice. *Development* **129**, 4797–4806 (2002).
- Joukov, V. et al. Proteolytic processing regulates receptor specificity and activity of VEGF-C. *EMBO J.* **16**, 3898–3911 (1997).
- Cao, Y. et al. Vascular endothelial growth factor C induces angiogenesis in vivo. *Proc. Natl Acad. Sci. USA* **95**, 14389–14392 (1998).
- Stacker, S. A. et al. Biosynthesis of vascular endothelial growth factor-D involves proteolytic processing which generates non-covalent homodimers. *J. Biol. Chem.* **274**, 32127–32136 (1999).
- Rissanen, T. T. et al. VEGF-D is the strongest angiogenic and lymphangiogenic effector among VEGFs delivered into skeletal muscle via adenoviruses. *Circ. Res.* **92**, 1098–1106 (2003).
- Nagy, J. A. et al. Vascular permeability factor/vascular endothelial growth factor induces lymphangiogenesis as well as angiogenesis. *J. Exp. Med.* **196**, 1497–1506 (2002).
- Karkkainen, M. J. et al. Vascular endothelial growth factor C is required for sprouting of the first lymphatic vessels from embryonic veins. *Nature Immunol.* **5**, 74–80 (2004).
- Gannon, G. et al. Overexpression of vascular endothelial growth factor-A165 enhances tumor angiogenesis but not metastasis during beta-cell carcinogenesis. *Cancer Res.* **62**, 603–608 (2002).
- Oh, S. J. et al. VEGF and VEGF-C: specific induction of angiogenesis and lymphangiogenesis in the differentiated avian chorioallantoic membrane. *Dev. Biol.* **188**, 96–109 (1997).
- Cursiefen, C. et al. VEGF-A stimulates lymphangiogenesis and hemangiogenesis in inflammatory neovascularization via macrophage recruitment. *J. Clin. Invest.* **113**, 1040–1050 (2004).
- Baluk, P. et al. Pathogenesis of persistent lymphatic vessel hyperplasia in chronic airway inflammation. *J. Clin. Invest.* **115**, 247–257 (2005).
- Wang, J. F., Zhang, X. F. & Grooman, J. E. Stimulation of beta 1 integrin induces tyrosine phosphorylation of vascular endothelial growth factor receptor-3 and modulates cell migration. *J. Biol. Chem.* **276**, 41950–41957 (2001).
- Vlahakis, N. E., Young, B. A., Atakilit, A. & Sheppard, D. The lymphangiogenic vascular endothelial growth factors VEGF-C and -D are ligands for the integrin alpha9beta1. *J. Biol. Chem.* **280**, 4544–4552 (2005).
- Huang, X. Z. et al. Fatal bilateral chylothorax in mice lacking the integrin alpha9beta1. *Mol. Cell Biol.* **20**, 5208–5215 (2000).
- Zhang, X. et al. KSHV activation of VEGFR-3 alters endothelial function and enhances infection. *J. Biol. Chem.* **280**, 26216–26224 (2005).
- Kubo, H. et al. Blockade of vascular endothelial growth factor receptor-3 signaling inhibits fibroblast growth factor-2-induced lymphangiogenesis in mouse cornea. *Proc. Natl Acad. Sci. USA* **99**, 8868–8873 (2002).
- Cao, R. et al. PDGF-BB induces intratumoral lymphangiogenesis and promotes tumour metastasis. *Cancer Cell* **6**, 333–345 (2004).
- Kajiya, K., Hirakawa, S., Ma, B., Drinnenberg, I. & Detmar, M. Hepatocyte growth factor promotes lymphatic vessel formation and function. *EMBO J.* **24**, 2885–2895 (2005).
- Sabin, F. R. The lymphatic system in human embryos, with a consideration of the morphology of the system as a whole. *Am. J. Anat.* **9**, 43–91 (1909).
- Wigle, J. T. & Oliver, G. Prox1 function is required for the development of the murine lymphatic system. *Cell* **98**, 769–778 (1999).
- Wigle, J. T. et al. An essential role for Prox1 in the induction of the lymphatic endothelial cell phenotype. *EMBO J.* **21**, 1505–1513 (2002).
- Hong, Y. K. et al. Prox1 is a master control gene in the program specifying lymphatic endothelial cell fate. *Dev. Dyn.* **225**, 351–357 (2002).
- Petrova, T. V. et al. Lymphatic endothelial reprogramming of vascular endothelial cells by the Prox-1 homeobox transcription factor. *EMBO J.* **21**, 4593–4599 (2002).
- Harvey, N. L. et al. Lymphatic vascular defects promoted by Prox1 haploinsufficiency cause adult-onset obesity. *Nature Genet.* **37**, 1072–1081 (2005).
- Baldwin, M. E. et al. Vascular endothelial growth factor D is dispensable for development of the lymphatic system. *Mol. Cell Biol.* **25**, 2441–2449 (2005).
- Dumont, D. J. et al. Cardiovascular failure in mouse embryos deficient in VEGF receptor-3. *Science* **282**, 946–949 (1998).
- Karkkainen, M. J. et al. Missense mutations interfere with VEGFR-3 signaling in primary lymphoedema. *Nature Genet.* **25**, 153–159 (2000).
- Jackson, D. G. Biology of the lymphatic marker LYVE-1 and applications in research into lymphatic trafficking and lymphangiogenesis. *APMIS* **112**, 526–538 (2004).
- Mäkinen, T. et al. PDZ interaction site in ephrinB2 is required for the remodeling of lymphatic vasculature. *Genes Dev.* **19**, 397–410 (2005).
- Jeltsch, M., Tammela, T., Alitalo, K. & Wiltling, J. Genesis and pathogenesis of lymphatic vessels. *Cell Tissue Res.* **314**, 69–84 (2003).
- Aalami, O. O., Allen, D. B. & Organ, C. H. Jr. Chylous ascites: a collective review. *Surgery* **128**, 761–778 (2000).
- Abtahian, F. et al. Regulation of blood and lymphatic vascular separation by signaling proteins SLP-76 and Syk. *Science* **299**, 247–251 (2003).
- Saharinen, P. et al. Multiple angiopoietin recombinant proteins activate the Tie1 receptor tyrosine kinase and promote its interaction with Tie2. *J. Cell Biol.* **169**, 239–243 (2005).
- Gale, N. W. et al. Angiopoietin-2 is required for postnatal angiogenesis and lymphatic patterning, and only the latter role is rescued by Angiopoietin-1. *Dev. Cell* **3**, 411–423 (2002).
- Morisada, T. et al. Angiopoietin-1 promotes LYVE-1-positive lymphatic vessel formation. *Blood* **105**, 4649–4656 (2005).
- Tammela, T. et al. Angiopoietin-1 promotes lymphatic sprouting and hyperplasia. *Blood* **105**, 4642–4648 (2005).
- Adams, R. H. & Klein, R. Eph receptors and ephrin ligands: essential mediators of vascular development. *Trends Cardiovasc. Med.* **10**, 183–188 (2000).
- Petrova, T. V. et al. Defective valves and abnormal mural cell recruitment underlie lymphatic vascular failure in lymphedema distichiasis. *Nature Med.* **10**, 974–981 (2004).
- Dagenais, S. L. et al. Foxc2 is expressed in developing lymphatic vessels and other tissues associated with lymphedema-distichiasis syndrome. *Gene Expr. Patterns* **4**, 611–619 (2004).
- Ferrell, R. E. Research perspectives in inherited lymphatic disease. *Ann. N.Y. Acad. Sci.* **979**, 39–51 (2002).
- Brice, G. et al. Analysis of the phenotypic abnormalities in lymphoedema-distichiasis syndrome in 74 patients with FOXC2 mutations or linkage to 16q24. *J. Med. Genet.* **39**, 478–483 (2002).
- Schacht, V. et al. Tlalpa/podoplanin deficiency disrupts normal lymphatic vasculature formation and causes lymphedema. *EMBO J.* **22**, 3546–3556 (2003).
- Kriehuber, E. et al. Isolation and characterization of dermal lymphatic and blood endothelial cells reveal stable and functionally specialized cell lineages. *J. Exp. Med.* **194**, 797–808 (2001).
- Mäkinen, T. et al. Isolated lymphatic endothelial cells transduce growth, survival and migratory signals via the VEGF-C/D receptor VEGFR-3. *EMBO J.* **20**, 4762–4773 (2001).
- Podgrabska, S. et al. Molecular characterization of lymphatic endothelial cells. *Proc. Natl Acad. Sci. USA* **99**, 16069–16074 (2002).
- Hirakawa, S. et al. Identification of vascular lineage-specific genes by transcriptional profiling of isolated blood vascular and lymphatic endothelial cells. *Am. J. Pathol.* **162**, 575–586 (2003).
- Rockson, S. G. Lymphedema. *Am. J. Med.* **110**, 288–295 (2001).
- Yla-Herttuala, S. & Alitalo, K. Gene transfer as a tool to induce therapeutic vascular growth. *Nature Med.* **9**, 694–701 (2003).
- Saariisto, A. et al. Lymphangiogenic gene therapy with minimal blood vascular side effects. *J. Exp. Med.* **196**, 719–730 (2002).
- Thurston, G. et al. Angiopoietin-1 protects the adult vasculature against plasma leakage. *Nature Med.* **6**, 460–463 (2000).
- Weis, S. & Cheresh, D. A. Pathophysiological consequences of VEGF-induced vascular permeability. *Nature* **65**, 9789–9798 (2005).
- Pepper, M. S. Lymphangiogenesis and tumor metastasis: myth or reality? *Clin. Cancer Res.* **7**, 462–468 (2001).
- Stacker, S. A., Achen, M. G., Jussila, L., Baldwin, M. E. & Alitalo, K. Lymphangiogenesis and cancer metastasis. *Nature Rev. Cancer* **2**, 573–583 (2002).

63. He, Y. *et al.* Suppression of tumor lymphangiogenesis and lymph node metastasis by blocking vascular endothelial growth factor receptor 3 signaling. *J. Natl Cancer Inst.* **94**, 819–825 (2002).
64. Skobe, M. *et al.* Induction of tumor lymphangiogenesis by VEGF-C promotes breast cancer metastasis. *Nature Med.* **7**, 192–198 (2001).
65. Mandriota, S. J. *et al.* Vascular endothelial growth factor-C-mediated lymphangiogenesis promotes tumour metastasis. *EMBO J.* **20**, 672–682 (2001).
66. Stacker, S. A. *et al.* Vascular endothelial growth factor-D promotes the metastatic spread of cancer via the lymphatics. *Nature Med.* **7**, 186–191 (2001).
67. Karpanen, T. *et al.* Vascular endothelial growth factor C promotes tumor lymphangiogenesis and intralymphatic tumor growth. *Cancer Res.* **61**, 1786–1790 (2001).
68. Valtola, R. *et al.* VEGFR-3 and its ligand VEGF-C are associated with angiogenesis in breast cancer. *Am. J. Pathol.* **154**, 1381–1390 (1999).
69. Krishnan, J. *et al.* Differential *in vivo* and *in vitro* expression of vascular endothelial growth factor (VEGF)-C and VEGF-D in tumors and its relationship to lymphatic metastasis in immunocompetent rats. *Cancer Res.* **63**, 713–722 (2003).
70. Dadras, S. S. *et al.* Tumor lymphangiogenesis: a novel prognostic indicator for cutaneous melanoma metastasis and survival. *Am. J. Pathol.* **162**, 1951–1960 (2003).
71. Maula, S. M. *et al.* Intratumoral lymphatics are essential for the metastatic spread and prognosis in squamous cell carcinomas of the head and neck region. *Cancer Res.* **63**, 1920–1926 (2003).
72. Padera, T. P. *et al.* Pathology: cancer cells compress intratumour vessels. *Nature* **427**, 695 (2004).
73. Wong, S. Y. *et al.* Tumor-secreted VEGF-C is necessary for prostate cancer lymphangiogenesis, but lymphangiogenesis is unnecessary for lymph node metastasis. *Cancer Res.* **65**, 9789–9798.
74. Padera, T. P. *et al.* Lymphatic metastasis in the absence of functional intratumor lymphatics. *Science* **296**, 1883–1886 (2002).
75. He, Y. *et al.* Vascular endothelial cell growth factor receptor 3-mediated activation of lymphatic endothelium is crucial for tumor cell entry and spread via lymphatic vessels. *Cancer Res.* **65**, 4739–4746 (2005).
76. Laakkonen, P., Porkka, K., Hoffman, J. A. & Ruoslahti, E. A tumor-homing peptide with a targeting specificity related to lymphatic vessels. *Nature Med.* **8**, 751–755 (2002).
77. Zlotnik, A. Chemokines in neoplastic progression. *Semin. Cancer Biol.* **14**, 181–185 (2004).
78. Hirakawa, S. *et al.* VEGF-A induces tumor and sentinel lymph node lymphangiogenesis and promotes lymphatic metastasis. *J. Exp. Med.* **201**, 1089–1099 (2005).
79. Pullinger, B. D. & Florey, H. W. Proliferation of lymphatics in inflammation. *J. Pathol. Bact.* **45**, 157–170 (1937).
80. Ristimäki, A., Narko, K., Enholm, B., Joukov, V. & Alitalo, K. Proinflammatory cytokines regulate expression of the lymphatic endothelial mitogen vascular endothelial growth factor-C. *J. Biol. Chem.* **273**, 8413–8418 (1998).
81. Saban, M. R. *et al.* Visualization of lymphatic vessels through NF- κ B activity. *Blood* **104**, 3228–3230 (2004).
82. Kerjaschki, D. *et al.* Lymphatic neoangiogenesis in human kidney transplants is associated with immunologically active lymphocytic infiltrates. *J. Am. Soc. Nephrol.* **15**, 603–612 (2004).
83. Hamrah, P., Chen, L., Zhang, Q. & Dana, M. R. Novel expression of vascular endothelial growth factor receptor (VEGFR)-3 and VEGF-C on corneal dendritic cells. *Am. J. Pathol.* **163**, 57–68 (2003).
84. Chen, L. *et al.* Vascular endothelial growth factor receptor-3 mediates induction of corneal alloimmunity. *Nature Med.* **10**, 813–815 (2004).
85. Ohl, L. *et al.* CCR7 governs skin dendritic cell migration under inflammatory and steady-state conditions. *Immunity* **21**, 279–288 (2004).
86. Irla, H. *et al.* Mannose receptor is a novel ligand for L-selectin and mediates lymphocyte binding to lymphatic endothelium. *J. Exp. Med.* **194**, 1033–1041 (2001).
87. Salmi, M., Koskinen, K., Henttinen, T., Elima, K. & Jalkanen, S. CLEVER-1 mediates lymphocyte transmigration through vascular and lymphatic endothelium. *Blood* **104**, 3849–3857 (2004).
88. Nibbs, R. J. *et al.* The beta-chemokine receptor D6 is expressed by lymphatic endothelium and a subset of vascular tumors. *Am. J. Pathol.* **158**, 867–877 (2001).
89. Kriederman, B. M. *et al.* FOXC2 haploinsufficient mice are a model for human autosomal dominant lymphedema-distichiasis syndrome. *Hum. Mol. Genet.* **12**, 1179–1185 (2003).
90. Ayadi, A. *et al.* Net-targeted mutant mice develop a vascular phenotype and up-regulate egr-1. *EMBO J.* **20**, 5139–5152 (2001).
91. Fruman, D. A. *et al.* Hypoglycaemia, liver necrosis and perinatal death in mice lacking all isoforms of phosphoinositide 3-kinase p85 α . *Nature Genet.* **26**, 379–382 (2000).
92. Pennisi, D. *et al.* Mutations in Sox18 underlie cardiovascular and hair follicle defects in ragged mice. *Nature Genet.* **24**, 434–437 (2000).
93. Gittenberger-De Groot, A. C. *et al.* Abnormal lymphatic development in trisomy 16 mouse embryos precedes nuchal edema. *Dev. Dyn.* **230**, 378–384 (2004).
94. Dixelius, J. *et al.* Ligand-induced vascular endothelial growth factor receptor-3 (VEGFR-3) heterodimerization with VEGFR-2 in primary lymphatic endothelial cells regulates tyrosine phosphorylation sites. *J. Biol. Chem.* **278**, 40973–40979 (2003).
95. Salameh, A., Galvagni, F., Bardelli, M., Bussolino, F. & Oliviero, S. Direct recruitment of CRK and GRB2 to VEGFR-3 induce proliferation, migration and survival of endothelial cells through the activation of ERK, AKT and JNK pathways. *Blood* **15**, 3423–3431 (2005).
96. Ny, A. *et al.* A genetic *Xenopus laevis* tadpole model to study lymphangiogenesis. *Nature Med.* **11**, 998–1004 (2005).
97. Schneider, M., Othman-Hassan, K., Christ, B. & Wilting, J. Lymphangioblasts in the avian wing bud. *Dev. Dyn.* **216**, 311–319 (1999).
98. Rajantie, I. *et al.* Adult bone marrow-derived cells recruited during angiogenesis comprise precursors for periendothelial vascular mural cells. *Blood* **104**, 2084–2086 (2004).
99. Wang, H. W. *et al.* Kaposi sarcoma herpesvirus-induced cellular reprogramming contributes to the lymphatic endothelial gene expression in Kaposi sarcoma. *Nature Genet.* **36**, 687–693 (2004).
100. Hong, Y. K. *et al.* Lymphatic reprogramming of blood vascular endothelium by Kaposi sarcoma-associated herpesvirus. *Nature Genet.* **36**, 683–685 (2004).

Acknowledgements We extend our gratitude to the many colleagues who have contributed to the field, but whose work could not be cited here owing to space limitations. We thank C. Norrmén for providing confocal image of lymphatic vessels, A. Parsons for assistance in editing, and H. Schmidt for the drawing of figures. The work in the authors' laboratories is supported by the US National Institutes of Health, the European Union, the Finnish Academy, the Sigrid Juselius Foundation and the Finnish Cancer Organizations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to K.A. (Kari.Alitalo@Helsinki.FI).

From angiogenesis to neuropathology

David A. Greenberg¹ & Kunlin Jin¹

Angiogenesis — the growth of new blood vessels — is a crucial force for shaping the nervous system and protecting it from disease. Recent advances have improved our understanding of how the brain and other tissues grow new blood vessels under normal and pathological conditions. Angiogenesis factors, especially vascular endothelial growth factor, are now known to have roles in the birth of new neurons (neurogenesis), the prevention or mitigation of neuronal injury (neuroprotection), and the pathogenesis of stroke, Alzheimer's disease and motor neuron disease. As our understanding of pathophysiology grows, these developments may point the way towards new molecular and cell-based therapies.

Diseases of the brain's vasculature cause more deaths than circulatory disorders of any other organ except the heart. Moreover, survivors of cerebral vascular disease are often severely incapacitated, reflecting the key role of the brain in determining who we are and what we can do. Brain cells die rapidly when deprived of their blood supply, and their high degree of specialization makes it difficult for surviving cells to assume their functions. All these considerations make it especially important to understand the relationship between the brain and its vasculature, both in development and in disease.

An examination of this relationship is timely, because the molecular mechanisms involved in angiogenesis and in the pathogenesis of neurological diseases are being deciphered at an unprecedented rate. Nevertheless, these insights have had little impact on the plight of the patient with stroke, lying on a hospital trolley in the emergency room, awaiting treatment that, for the most part, does not exist.

Angiogenesis and neurogenesis are prominent features of neurological disease, either as pathophysiological factors or as responses to injury. One common thread that connects angiogenesis, neurogenesis and pathogenesis is vascular endothelial growth factor (VEGF, or VEGF-A), which was identified on the basis of its vascular effects, but has since been recognized as an important signalling molecule in the nervous system as well. Recent insights into the role of VEGF in a variety of neurological disorders, including stroke and motor neuron disease, suggest that VEGF or its downstream effectors may be promising therapeutic targets in these diseases.

This review will trace the thread of recent advances that could connect knowledge of how the cerebral vasculature develops to new therapeutic insights into neurological disease. The topics addressed include angiogenesis in the normal and diseased brain, links between angiogenesis and brain development or protection, and the role of angiogenesis factors such as VEGF in pathology of the central nervous system (CNS).

Cerebral angiogenesis

The CNS acquires its vascular supply through angiogenesis, or the sprouting of new capillaries from pre-existing vessels, rather than vasculogenesis, in which angioblasts differentiate into endothelial cells to form blood vessels *de novo*^{1,2}. Early in embryogenesis, vessels from the pia mater invade the brain and converge centripetally towards the ventricles; some deeply penetrating vessels give off second-order branches that surround the ventricles, or extend centrifugally back toward the

pia (Fig. 1). This produces a marginally perfused border zone, or watershed, between the centripetal and centrifugal vessels, which may render the ventricular zone particularly vulnerable to perinatal ischaemia (the reduction of blood supply), as in cerebral palsy.

By adulthood, the human cerebral circulation sustains blood flow of about 50 ml per 100 g per min, and receives roughly 20% of cardiac output, or about ten times the flow that would be expected on the basis of brain mass. The cerebral circulation is also distinctive in that its capillaries harbour a blood–brain barrier, produced by tight junctions between endothelial cells, which impedes the entry of many molecules from the blood into the CNS³. The blood–brain barrier is disrupted in a variety of diseases, including stroke and brain tumours, and this leads

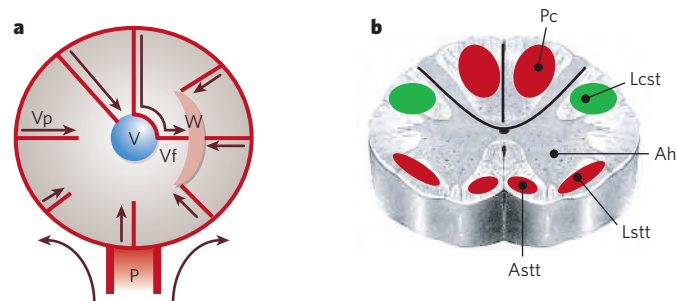


Figure 1 | Vascularization of the central nervous system. **a**, Pattern of developmental cerebral angiogenesis. Vessels from the pia mater (P) invade the CNS, where they begin to grow in a ventriculopetal (Vp) direction. Some reach the ventricles (V), and then give off branches that surround them. These, in turn, give rise to ventriculofugal (Vf) branches, which grow back towards the pia. The border zone between ventriculopetal and ventriculofugal vessels is a watershed zone (W), which may be susceptible to ischaemic injury if systemic perfusion pressure decreases. **b**, Vascular supply of the adult spinal cord. The spinal cord, shown in horizontal section, is supplied by the anterior spinal artery and paired posterior spinal arteries; the borders between these three vascular territories are shown in black. The anterior spinal artery supplies the anterior part of the spinal cord bilaterally, including both lateral corticospinal tracts (Lcst), which are descending motor pathways that project onto motor neurons in the anterior horns (Ah), as well as the lateral (Lstt) and anterior (Astt) spinothalamic tracts, which convey pain and temperature (and to some extent touch and pressure) sensation to the brain. Each posterior spinal artery supplies a posterior column (Pc), which also transmits touch and pressure sensation to the brain.

¹Buck Institute for Age Research, 8001 Redwood Boulevard, Novato, California 94945, USA.

to cerebral oedema, which may contribute to increased intracranial pressure and result in death.

Several angiogenesis factors and downstream signalling pathways have been implicated in cerebral angiogenesis. The most prominent of these is VEGF^{4,5}, which is expressed in the brain and is induced by hypoxia through the transcription factor HIF-1 (hypoxia-inducible factor-1)⁶. Interestingly, however, in mice with deletion of brain HIF-1, VEGF expression is preserved and outcome after global cerebral ischaemia is improved⁷. The tyrosine kinase receptors and downstream protein kinases implicated in VEGF signalling in the CNS have been essentially identical to those described previously in non-neural tissues, and include VEGFR-2 and to some extent VEGFR-1 receptors, as well as mitogen-activated protein kinases (MAPKs), phosphatidylinositol 3-kinase (PI3K)/Akt, and Rho/Rac^{8,9}. These and related angiogenesis signalling pathways are reviewed in more detail in the review in this issue by Coultas, Chawengsaksophak and Rossant, p. 937.

Angiogenesis in brain diseases

Although the role of angiogenesis in the CNS is best characterized in development¹⁰, cerebral angiogenesis also accompanies tissue grafting^{11–13} and tumour growth^{14,15}. Neural grafts into the brain become vascularized, with the source of vessels (donor or recipient) and the time course of vascularization dependent on whether solid tissue or dissociated cells are transplanted, whether they are placed in the ventricles or brain parenchyma, and whether allografts or xenografts are used. Brain tumour angiogenesis is another example of the ability of the adult brain to produce new blood vessels, but presents serious clinical problems¹⁶, because it allows tumours to grow to sizes that, within the rigid confines of the skull, can cause brain herniation and death. The fact that tumour vessels lack a blood–brain barrier, and are consequently associated with peritumoural oedema, compounds this problem, as does the proclivity of these abnormal vessels in certain tumours — including glioblastoma multiforme and metastatic melanoma, choriocarcinoma, renal cell carcinoma and bronchogenic carcinoma — to bleed. However, their abnormal vasculature also renders tumours vulnerable to anti-angiogenic therapy, potential approaches to which include VEGFR-2 tyrosine kinase inhibitors or antibodies, soluble decoy receptors that sequester VEGF, or ‘hunter-killer’ peptides addressed specifically to tumour vessels¹⁷. By attacking the tumour vasculature rather than the tumour itself, these strategies may partly sidestep current limitations to cancer treatment, such as drug resistance and toxic effects on rapidly proliferating non-neoplastic tissue. Alternatively, tumour vessels may also have the capacity to adapt to anti-angiogenic treatment (see the review in this issue by Ferrera and Kerbel, p. 967).

Hypoxia or ischaemia can induce angiogenesis in most tissues, including the CNS. In adult humans, pure hypoxia is rare, but cerebral ischaemia is not. Transient global ischaemia, which complicates cardiac arrest, causes death of intrinsically susceptible neurons, such as those in the CA1 region of the hippocampus, and of cells in selectively vulnerable areas, such as the vascular watershed or border zones of the brain and spinal cord. The more common cause of vascular brain injury is focal cerebral ischaemia, which may be transient or permanent. It is usually attributed to atherosclerotic occlusion or embolism within an artery. Focal cerebral ischaemia of sufficient severity and duration to result in death of tissue, or infarction, produces stroke, which is characterized by a persistent disturbance of neurological function associated with damage to a discrete area of the brain. The localized nature of brain injury in stroke accounts for the stereotypical syndromes it causes. If blood flow is restored promptly, usually within 30 minutes, there may be no permanent impairment, and a transient ischaemic attack is diagnosed. In fact, transient ischaemia may even help to protect the brain from subsequent ischaemic injury, at least for a brief interval. This is known as ischaemic tolerance or preconditioning¹⁸, and measures that recapitulate the molecular events in preconditioning may have therapeutic potential in stroke. But current treatment for stroke is poor. Aspirin and other antiplatelet drugs have

some preventive value, as does carotid endarterectomy (removal of a clot from a partially occluded carotid artery), and thrombolytic drugs such as tissue plasminogen activator can relieve cerebrovascular occlusions, but only if administered within the first few hours after stroke. The ultimate role of angioplasty (temporary widening of the artery with a catheter) and stenting (inserting a permanent support of fine wire mesh) in disorders of the cerebral circulation is not yet clear. So additional, more effective stroke treatment is needed, making efforts to improve cerebral angiogenesis attractive.

Autopsy studies show that brain ischaemia stimulates angiogenesis¹⁹, and animal models of stroke²⁰ suggest that this becomes evident within 1–2 weeks. New vessel growth is most pronounced in the ischaemic penumbra, where blood flow is reduced but not absent, and where small changes in perfusion might make the difference between cell death and survival. Clinical studies suggest that rescuing tissue within the penumbra improves recovery, but it is unclear whether ischaemia-induced angiogenesis modifies outcome from clinical stroke.

An increase in angiogenesis by VEGF in rats is associated with reduced neurological deficits²¹. Such a treatment strategy might be clinically useful, although the leakiness of newly produced vessels poses a risk of increasing swelling in the brain²². This problem might be obviated by treatment with a combination of VEGF and angiopoietins²³, or with HIF prolyl hydroxylase inhibitors, which raise HIF-1 levels and increase expression of several hypoxia-response proteins, including VEGF²⁴. The timescale of several days required to produce new vessels remains a problem in any angiogenesis-based approach to stroke, because neurons begin to die within minutes after their blood supply is interrupted. Nevertheless, there are patients in whom stroke is predictable or recurrent, such as those undergoing cardiac surgery or experiencing treatment-refractory transient ischaemic attacks, so that even therapy with a delayed effect could be useful. VEGF has been investigated clinically as a treatment for cardiac and limb ischaemia, but without definitive evidence of benefit (see p. 967). Whether this will also prove true for cerebral ischaemia is unclear. In animal models of stroke, delivering VEGF too early (within 1 hour of onset) worsens outcome by increasing brain oedema, whereas later administration (48 hours after onset) is beneficial²¹. Timing is also likely to be crucial in clinical settings. Cell-based approaches for stimulating angiogenesis in the ischaemic brain, such as angioblast transplantation, might also be feasible.

Some cerebral disorders are associated with haemorrhage, rather than ischaemia. Examples include ‘berry aneurysms’, which can bleed into the subarachnoid space, and vascular malformations, which cause haemorrhage into the brain itself. In most cases, the molecular aetiology is unknown, but in some instances, malformations of cerebral vessels are manifestations of genetic disorders that affect angiogenesis. Hereditary haemorrhagic telangiectasia, or Osler–Rendu–Weber disease, is caused by mutations in endoglin, a homodimeric membrane glycoprotein found primarily on endothelial cells²⁵. Endoglin expression is increased by cerebral hypoxia or ischaemia²⁶, consistent with a role in ischaemia-induced angiogenesis, but it is unclear how abnormal endoglin leads to the formation of weak, thin-walled blood vessels. Mutations in a protein of unknown function called KRIT1, which is expressed in endothelial cells undergoing tube formation during angiogenesis, are found in patients with cerebral cavernous malformations²⁵, which consist of dilated, tightly packed, capillary-like sinusoids. These genetic causes of cerebral haemorrhage are rare, but might be good candidates for therapy directed at gene replacement.

Adult neurogenesis

Cellular senescence or injury in the adult requires a capacity for generating new cells. This capacity resides in adult or tissue stem cells, which are self-renewing but less than totipotent, preferentially regenerating cell types indigenous to a particular tissue or organ. Adult cyto-genesis is best characterized in tissues with rapid cell turnover, such as bone marrow, intestine and skin, but is also found in the brain, where

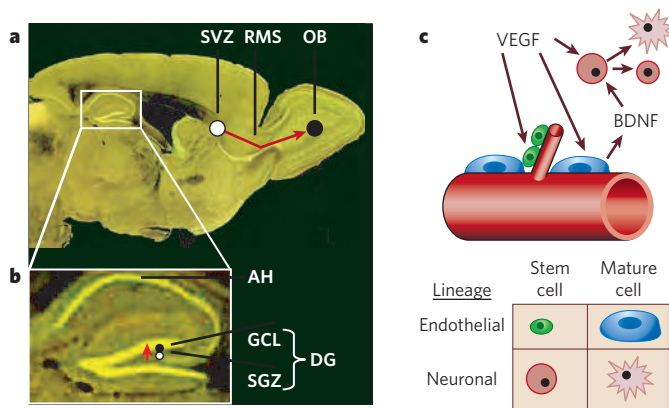


Figure 2 | Adult neurogenesis. **a**, Neurogenesis in the subventricular zone (SVZ). Neurons that arise in the rostral SVZ migrate by way of the rostral migratory stream (RMS) to the olfactory bulb (OB). **b**, Neurogenesis in the dentate gyrus (DG). Neurons arise in the dentate subgranular zone (SGZ) and migrate into the adjacent granule cell layer (GCL); AH, Ammon's horn of hippocampus. **c**, Concept of the 'vascular niche' and VEGF-induced neurogenesis. Neurogenesis is often observed close to vasculature, and certain growth factors, such as VEGF, stimulate both angiogenesis and neurogenesis. Possible explanations include parallel, independent effects of VEGF on endothelial and neuronal stem cells, or VEGF-induced production by endothelial cells of other growth factors, like BDNF, that act on neuronal stem cells to stimulate neurogenesis.

it occurs in at least two brain regions — the subventricular zone (SVZ) within the walls of the lateral ventricles, and the subgranular zone (SGZ) of the hippocampal dentate gyrus²⁷ (Fig. 2a, b). Here, neurogenesis is detected by the presence of cells that can be labelled with cell-proliferation markers such as bromodeoxyuridine (BrdU) and that also express markers of neuronal lineage, such as doublecortin (Dcx). Neurons that arise in the rodent SVZ migrate along a pathway called the rostral migratory stream to the olfactory bulb, where they replenish the supply of interneurons that undergo programmed cell death. In humans, however, the fate of neurons born in the SVZ is uncertain²⁸. Newborn neurons in the dentate subgranular zone migrate into the adjacent dentate granule cell layer, where they become mature, functional granule neurons²⁹, and may participate in memory and learning³⁰.

Brain injury caused by seizures^{31,32} or cerebral ischaemia^{33–35} can stimulate neurogenesis and direct the migration of newly produced neurons towards affected brain regions^{36–40}, possibly providing a mechanism for brain repair. Evidence for increased neurogenesis has also been reported in the brains of patients with Huntington's disease⁴¹ and Alzheimer's disease⁴², as well as in animal models of Alzheimer's disease⁴³, Huntington's disease⁴⁴ and perhaps Parkinson's disease^{45,46}. Liu and colleagues⁴⁷ have recently reported that in a transgenic mouse model of amyotrophic lateral sclerosis (ALS), neurogenesis in the spinal cord is increased, and neuronal precursors migrate from the central spinal canal to the site of motor neuron loss in the ventral horns. The fact that in all these diseases, the CNS responds with increased neuroproliferation and homing of newborn neurons to areas of injury suggests that it may be possible to make use of this innate propensity for therapeutic purposes. For example, if neuroproliferation could be stimulated by a drug or other intervention, might the brain or spinal cord's own repair programmes take over from that point?

Several growth factors promote adult neurogenesis²⁷, including fibroblast growth factor-2, epidermal growth factor, brain-derived neurotrophic factor, erythropoietin (BDNF), stem-cell factor, heparin-binding epidermal-growth-factor-like growth factor and VEGF. VEGF increases the incorporation of BrdU into cells expressing immature neuronal markers both in mouse cortical cultures *in vitro* and in SVZ and SGZ of adult rat brain *in vivo*^{48,49}. The VEGFR-2 receptor is implicated in each of these cases. In contrast, VEGF seems to stimu-

late the growth of astrocytes by activating VEGFR-1 (ref. 50). Growth factors seem to be good candidates for stimulating therapeutic neurogenesis, especially if they can be given by relatively non-invasive routes, such as through the nose⁵¹. In this manner, it might also be possible to avoid adverse effects outside the brain. On the other hand, we know relatively little about how different growth factors target different subpopulations of neuronal precursors or direct these cells towards different phenotypic fates. Certain growth factors may be better suited for replacing neurons in some disorders than in others. For example, VEGF, which is both angiogenic and neurogenic, might be a good fit in stroke, but not in pure neuronal degenerations.

The functional significance of adult neurogenesis, and of VEGF-induced adult neurogenesis in particular, is not fully understood. An interesting example of the latter, however, is hippocampal neurogenesis induced by experiences such as environmental enrichment and learning. Rats raised in a physically enriched environment or trained in a Morris water maze task show increased hippocampal expression of VEGF⁵². Moreover, increasing hippocampal VEGF expression and neurogenesis, by intracerebral administration of a VEGF-expressing adeno-associated viral vector, improves hippocampus-dependent associative and spatial learning. As in most other studies of VEGF and neurogenesis, this effect involves VEGFR-2. Conversely, small hairpin RNAs that knock down hippocampal VEGF expression block the induction of neurogenesis by environmental enrichment.

In a 1978 review on hematopoietic stem cells⁵³, Schofield coined the term 'stem-cell niche' to describe the extracellular environment that maintains a stem cell in its undifferentiated, proliferating state. In the specific case of neuronal stem cells in the hippocampus, Palmer and colleagues have proposed that vascular elements are an essential feature of this niche⁵⁴ (Fig. 2c). They observed that clusters of proliferating cells in the dentate SGZ contain cells expressing neuronal, glial and endothelial markers, and tend to occur close to capillaries. In addition, they found high levels of VEGF and VEGFR-2 in these regions, consistent with a role for VEGF in coupling angiogenesis with neurogenesis. This would be consistent with a generalized link between vascular signalling and organogenesis⁹. Alternatively, cells of different lineages could be associated with the vasculature because they respond to a haematogenously transmitted common signal. Goldman and colleagues proposed a specific mechanistic basis for the vascular niche in the adult songbird brain⁵⁵. They found evidence for a signalling sequence in which testosterone upregulates VEGF and VEGFR-2 in the higher vocal centre, leading to angiogenesis and endothelial production of BDNF, which in turn acts on neuronal progenitors to stimulate neurogenesis. Temple and co-workers reported that cultured neuronal stem cells from embryonic mouse cortex proliferate more rapidly in the presence of a mouse brain endothelial cell line⁵⁶. However, the fact that VEGF promotes neurogenesis in vessel-free neuronal cultures *in vitro*⁴⁸ implies that a vascular niche may not always be required for neurogenesis to occur. Evidence for the converse phenomenon — an effect of neuronal VEGF on vascular development — has also been presented^{57,58}. Finally, some evidence suggests that whereas VEGF is required for normal developmental neurogenesis, neuronal VEGFR-2 expression is not⁵⁹.

Neuroprotection and neurodegeneration

The earliest described direct neuronal actions of VEGF were neurotrophic effects. Sondell *et al.*⁶⁰ reported that, in cultured superior cervical and dorsal root ganglion neurons, VEGF promotes both axonal outgrowth and cell survival by a VEGFR-2-dependent mechanism, and Silverman *et al.*⁶¹ showed that VEGF improves neuronal survival in organotypic midbrain explant cultures. Subsequent studies demonstrated that VEGF caused increases in the number⁶² or length⁶³ of neurites in cultured neurons or cortical explants, which involved signalling through VEGFR-2, MAPK and PI3K/Akt (Fig. 3). Rho/Rac signalling has also been implicated in VEGF-induced neurite outgrowth⁶⁴.

In addition to these trophic, or growth-promoting, effects, VEGF can also protect neurons from a range of insults. For example, VEGF reduces death of immortalized hippocampal neurons subjected to

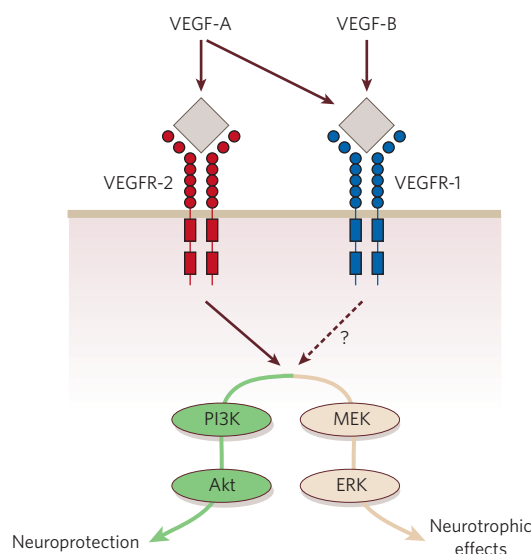


Figure 3 | Mechanisms of direct neuronal effects of VEGF. VEGF acts through multiple tyrosine kinase receptors and associated downstream signalling pathways. In neurons, the best characterized of these are VEGFR-2 and the PI(3)K/Akt and MEK/ERK protein kinase pathways.

serum withdrawal or hypoxia *in vitro*, by activating VEGFR-2, PI3K/Akt, and the transcription factor NF- κ B^{65,66}. VEGF also protects cultured cerebral cortical neurons from hypoxia through similar mechanisms⁶⁶, as well as by decreasing activation of caspase-3 (ref. 67). Finally, VEGF protects cultured hippocampal neurons from glutamate⁶⁸ and *N*-methyl-D-aspartate⁶⁹ toxicity.

VEGF modifies acute and chronic neurodegenerative processes through effects on both blood vessels and neurons, and probably also glia. Cerebral ischaemia stimulates VEGF expression^{70–73} and VEGF, in turn, promotes cerebral angiogenesis^{74,75}. Topical application of VEGF reduces brain infarct size⁷⁶ and intravenous VEGF improves neurological outcome from ischaemia²¹, whereas intraventricular administration of an anti-VEGF antibody increases infarct size⁷⁷. But

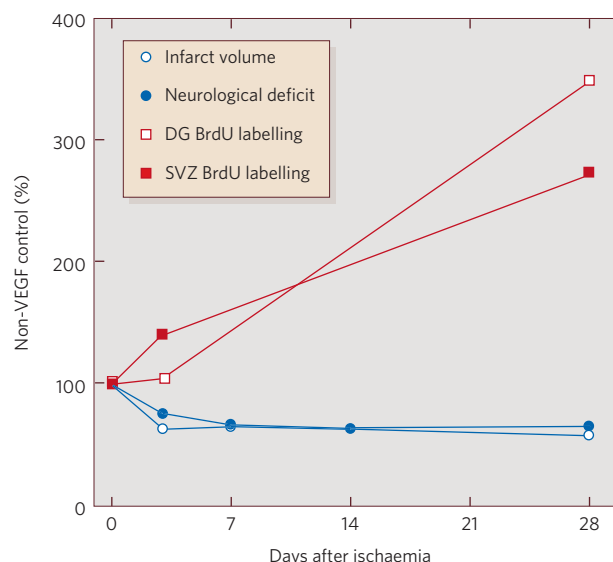


Figure 4 | VEGF in stroke. VEGF, given by intraventricular infusion on days 1–3 after middle cerebral artery occlusion in rats, reduced infarct volume and neurological deficits with maximal effects that were already evident by 3 days post-ischaemia and persisted for at least 28 days. In contrast, VEGF-induced neurogenesis in DG and SVZ was apparent at 28 days, but not at 3 days. Thus, although early post-ischaemic VEGF increased neurogenesis and improved outcome, the two effects do not seem to be causally related (data from ref. 78).

intraperitoneal administration of a fusion protein that sequesters VEGF reduced infarct size in another study²², indicating that the effects of VEGF in cerebral ischaemia are not always beneficial.

When infused into the lateral ventricles on days 1–3 after occlusion of the middle cerebral artery in rats, VEGF reduced infarct size and improved neurological outcome, with a maximal effect at 3 days (Fig. 4); improved neurogenesis, which was evident between 3 and 28 days; and stimulated angiogenesis in the ischaemic penumbra⁷⁸. This suggests that direct neuroprotection may reduce ischaemic brain damage in the acute phase, whereas neurogenesis and angiogenesis may be involved in long-term repair. Like VEGF, VEGF-B also seems to exert a neuroprotective effect in cerebral ischaemia. Accordingly, VEGF-B protects cultured cerebral cortical neurons from hypoxia, whereas infarct volume is increased and neurological function is more impaired in VEGF-B-knockout than in wild-type mice⁷⁹.

Some studies have drawn a connection between VEGF and degenerative brain diseases. Kalaria *et al.*⁸⁰ reported increased VEGF immunoreactivity in clusters of reactive astrocytes from brains of patients with Alzheimer's disease. Single-nucleotide polymorphisms in the promoter regions of the VEGF gene have also been associated with increased risk of Alzheimer's disease⁸¹.

VEGF may also have a role in diseases of the peripheral nervous system. Thus, intramuscular administration of plasmid DNA encoding VEGF improved nerve function in rats with diabetic (streptozotocin-induced) polyneuropathy⁸², and intramuscular delivery of naked VEGF DNA had a similar effect in rabbits with ischaemic neuropathy⁸³. Recently, Fink and colleagues⁸⁴ administered a replication-defective herpes simplex virus vector expressing VEGF by the intramuscular route to mice with streptozotocin neuropathy and found preservation of nerve fibres, increased sensory nerve amplitudes and thermal pain perception, improved autonomic function and greater nerve vascularity.

Another exciting development implicating VEGF in neurodegenerative disorders has been its association with ALS. A recent biography⁸⁵ details the manner in which this disease led to the premature retirement and death of the famed New York Yankee baseball player Henry Louis (Lou) Gehrig. ALS causes loss of lower motor neurons in the spinal cord and brainstem, as well as upper motor neurons in the motor cortex. Although most cases are sporadic,

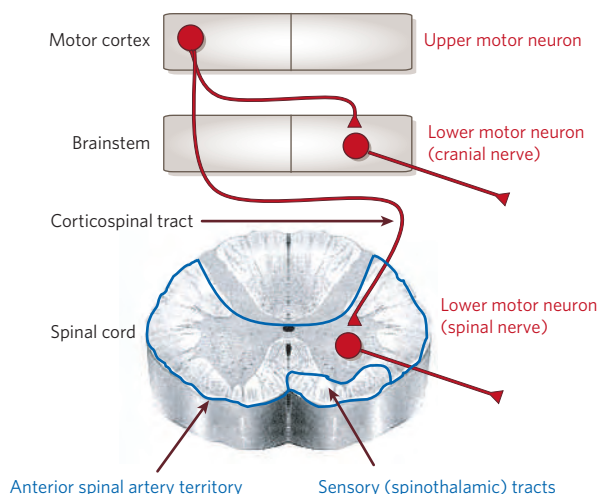


Figure 5 | VEGF and motor neuron disease. Variations in the VEGF gene have been implicated in experimental and clinical ALS, but whether this implies a hypoxic or ischaemic aetiology for ALS is unclear. ALS affects upper and lower motor neurons (red), but not sensory neurons or tracts, whereas spinal cord ischaemia, usually caused by occlusion of the anterior spinal artery, affects both motor and sensory (spinothalamic) pathways (blue).

some result from mutations in Cu/Zn superoxide dismutase (SOD1; ref. 85), the guanine nucleotide exchange factor alsin (ALS2; refs 87, 88), the DNA/RNA α -helicase senataxin⁹¹, or the vesicle trafficking protein VAPB⁹⁰. In 2001, Carmeliet and colleagues found that mice with a mutation in a site that confers hypoxia-responsiveness on VEGF (*VEGF*^{6/6}) developed a disorder manifested by limb weakness and

neurogenic muscle atrophy, associated with electrophysiological signs of muscle denervation and reinnervation, and loss of motor neurons from the anterior horn of the spinal cord and brainstem motor nuclei⁹¹. Although the site of the mutation suggested that affected mice might lose anterior horn cells because they failed to respond to spinal cord hypoxia or ischaemia, there is little other evidence for a hypoxic or ischaemic aetiology in clinical motor neuron disease. Ischaemic lesions of the spinal cord usually result from occlusion of the anterior spinal artery, affecting not only motor (anterior horns and corticospinal tracts) but also sensory (spinothalamic) pathways, or from a drop in blood pressure, producing both motor and sensory involvement due to infarction in the mid-thoracic watershed region (Fig. 5). This differs markedly from the distribution of involvement in motor neuron disease, which specifically involves motor neurons in the anterior horns and the corticospinal tracts. In addition, neither brainstem motor nuclei affected in ALS nor upper motor neurons in cerebral cortex are especially vulnerable to hypoxia. Finally, VEGF improves the survival of cultured spinal cord motor neurons from wild-type, as well as *VEGF*^{6/6} mice⁹¹. Its protective effect on motor neurons, like most of its other neuronal effects, was mediated by VEGFR-2, and also seemed to involve a neuropilin 1 co-receptor.

In subsequent studies, genetic variations that decreased blood VEGF levels and increased ALS risk were detected in human subjects⁹². In addition, in transgenic mice that express a mutant human gene for ALS (SOD1G93A), pathology is reduced by intraperitoneal⁹³ or intraventricular⁹⁴ VEGF, or by intramuscular administration of a VEGF-expressing lentiviral vector that is transported retrogradely to motor neurons⁹⁵. These results and those described above for diabetic neuropathy highlight the potential of gene therapy approaches for delivering VEGF to neural tissues.

VEGF has also been implicated in another form of motor neuron disease, X-linked spinal and bulbar muscular atrophy or Kennedy's disease, which is caused by a polyglutamine expansion in the androgen receptor⁹⁶. Ellerby, La Spada and colleagues⁹⁷ produced transgenic mice expressing the expanded human androgen receptor, which caused a progressive disorder comprising muscle weakness, atrophy, weight loss and early death, and manifested histologically by loss of motor neurons from the anterior horns of the spinal cord and neurogenic (grouped) muscle fibre atrophy. Motor neurons from these mice showed reduced viability in cell culture, which was rescued by treatment with an isoform of VEGF (VEGF164). In this form of muscular atrophy, the expanded androgen receptor seems to interfere with transcriptional events that involve CREB-binding protein, which regulates *VEGF* gene expression. Thus, a disturbance in the neuronal effects of VEGF is observed in at least two genetically and phenotypically distinct forms of motor neuron disease.

Conclusion

How should research in this area now be directed to maximize its scientific and eventual clinical yield? Two approaches seem to merit particular focus efforts to revascularize the ischaemic brain and to make use of the neuroprotective and neuroregenerative actions of VEGF.

Revascularization strategies are limited by the fact that stroke is often unpredictable, and by the several days' time required for new blood vessels to form. On the other hand, stroke recurs within 3 years in up to 25% of patients, with the greatest risk occurring early⁹⁸. Consequently, targeting pro-angiogenic therapy to the first few months following a stroke might make sense. The risk of stroke is also high in the weeks following acute myocardial infarction, and in patients under-

going cardiac surgery, such as coronary artery bypass grafts. In these settings, measures to promote angiogenesis could reduce stroke risk and also benefit the ischaemic heart.

The first obvious opportunity for testing the potential of VEGF's neuroprotective action is in ALS, where preclinical evidence supports a role for VEGF, the natural history of the disease is forbidding, and existing treatment options are minimal. Nevertheless, the genetic rodent models of ALS that have been used in studies on VEGF are different enough from the usually sporadic human disease that critical questions remain. For example, how long will treatment be required and will upper and lower motor neurons respond to an equal extent? Diabetic and other polyneuropathies are also targets for neuron-directed VEGF therapy. The prospects for treatment of neurodegenerative diseases based on VEGF-induced neurogenesis (or neurogenesis in general) are even murkier. Evidence of functional benefit from neurogenesis, whether induced by injury or growth factor, is sparse⁹⁹ and awaits better techniques for selectively inhibiting neurogenesis to determine its impact. Nevertheless, it is encouraging that barely 15 years after the discovery of VEGF, a considerable amount has been learned about the molecular mechanisms involved in angiogenesis, and therapeutic applications for neurological disease may be on the horizon. ■

- Beck, L. Jr & D'Amore, P. A. Vascular development: cellular and molecular regulation. *FASEB J.* **11**, 365–373 (1997).
- Risau, W. Mechanisms of angiogenesis. *Nature* **386**, 671–674 (1997).
- Ballabh, P., Braun, A. & Nedergaard, M. The blood–brain barrier: an overview: structure, regulation, and clinical implications. *Neurobiol. Dis.* **16**, 1–13 (2004).
- Keck, P. J. et al. Vascular permeability factor, an endothelial cell mitogen related to PDGF. *Science* **246**, 1309–1312 (1989).
- Leung, D. W., Cachianes, G., Kuang, W.-J., Goeddel, D. V. & Ferrara, N. Vascular endothelial growth factor is a secreted angiogenic mitogen. *Science* **246**, 1306–1309 (1989).
- Goldberg, M. A. & Schneider, T. J. Similarities between the oxygen-sensing mechanisms regulating the expression of vascular endothelial growth factor and erythropoietin. *J. Biol. Chem.* **269**, 4355–4359 (1994).
- Helton, R. et al. Brain-specific knock-out of hypoxia-inducible factor-1 α reduces rather than increases hypoxic-ischemic damage. *J. Neurosci.* **25**, 4099–4107 (2005).
- Cross, M. J., Dixelius, J., Matsumoto, T. & Claesson-Welsh, L. VEGF-receptor signal transduction. *Trends Biochem. Sci.* **28**, 488–494 (2003).
- Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* **9**, 669–676 (2003).
- Ment, L. R. et al. Vascular endothelial growth factor mediates reactive angiogenesis in the postnatal developing brain. *Dev. Brain Res.* **100**, 52–61 (1997).
- Krum, J. M. & Rosenstein, J. M. Patterns of angiogenesis in neural transplant models: II. Fetal neocortical transplants. *J. Comp. Neurol.* **271**, 331–345 (1988).
- Broadwell, R. D. et al. Angiogenesis and the blood–brain barrier in solid and dissociated cell grafts within the CNS. *Prog. Brain Res.* **82**, 95–101 (1990).
- Geny, C. et al. Long-term delayed vascularization of human neural transplants to the rat brain. *J. Neurosci.* **14**, 7553–7562 (1994).
- Plate, K. H., Breier, G., Weich, H. A. & Risau, W. Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas *in vivo*. *Nature* **359**, 845–848 (1992).
- Kim, K. J. et al. Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth *in vivo*. *Nature* **362**, 841–844 (1993).
- Weis, S. M. & Cheresh, D. A. Pathophysiological consequences of VEGF-induced vascular permeability. *Nature* **437**, 497–504 (2005).
- Ellerby, H. M. et al. Anti-cancer activity of targeted pro-apoptotic peptides. *Nature Med.* **5**, 1032–1038 (1999).
- Dirnagl, U., Simon, R. P. & Hallenbeck, J. M. Ischemic tolerance and endogenous neuroprotection. *Trends Neurosci.* **26**, 248–254 (2003).
- Krupinski, J., Kaluza, J., Kumar, P., Kumar, S. & Wang, J. M. Role of angiogenesis in patients with cerebral ischemic stroke. *Stroke* **25**, 1794–1798 (1994).
- Chen, H. H., Chien, C.-H. & Liu, H. M. Correlation between angiogenesis and basic fibroblast growth factor expression in experimental brain infarct. *Stroke* **25**, 1651–1657 (1994).
- Zhang, Z. G. et al. VEGF enhances angiogenesis and promotes blood–brain barrier leakage in the ischemic brain. *J. Clin. Invest.* **106**, 829–838 (2000).
- van Bruggen, N. et al. VEGF antagonism reduces edema formation and tissue damage after ischemia/reperfusion injury in the mouse brain. *J. Clin. Invest.* **104**, 1613–1620 (1999).
- Thurston, G. et al. Angiopoietin-1 protects the adult vasculature against plasma leakage. *Nature Med.* **6**, 460–463 (2000).
- Sharp, F. R. & Bernaudin, M. HIF1 and oxygen sensing in the brain. *Nature Rev. Neurosci.* **5**, 437–448 (2004).
- Marchuk, D. A., Srinivasan, S., Squire, T. L. & Zawistowski, J. S. Vascular morphogenesis: tales of two syndromes. *Hum. Mol. Genet.* **12** Spec. No. 1, R97–R112 (2003).
- Zhu, Y. et al. Hypoxic induction of endoglin via mitogen-activated protein kinases in mouse brain microvascular endothelial cells. *Stroke* **34**, 2483–2488 (2003).
- Abrams, D. N., Koehl, M. & Le Moal, M. Adult neurogenesis: from precursors to network and physiology. *Physiol. Rev.* **85**, 523–569 (2005).
- Sasai, N. et al. Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration. *Nature* **427**, 740–744 (2004).
- van Praag, H. et al. Functional neurogenesis in the adult hippocampus. *Nature* **415**, 1030–1034 (2002).

30. Shors, T. J. *et al.* Neurogenesis in the adult is involved in the formation of trace memories. *Nature* **410**, 372–376 (2001).
31. Parent, J. M. *et al.* Dentate granule cell neurogenesis is increased by seizures and contributes to aberrant network reorganization in the adult rat hippocampus. *J. Neurosci.* **17**, 3727–3738 (1997).
32. Scharfman, H. E., Goodman, J. H. & Sollas, A. L. Granule-like neurons at the hilar/CA3 border after status epilepticus and their synchrony with area CA3 pyramidal cells: functional implications of seizure-induced neurogenesis. *J. Neurosci.* **20**, 6144–6158 (2000).
33. Liu, J., Solway, K., Messing, R. O. & Sharp, F. R. Increased neurogenesis in the dentate gyrus after transient global ischemia in gerbils. *J. Neurosci.* **18**, 7768–7778 (1998).
34. Gu, W., Brannstrom, T. & Wester, P. Cortical neurogenesis in adult rats after reversible photothrombotic stroke. *J. Cereb. Blood Flow Metab.* **20**, 1166–1173 (2000).
35. Jin, K. *et al.* Neurogenesis in dentate subgranular zone and rostral subventricular zone after focal cerebral ischemia in the rat. *Proc. Natl Acad. Sci. USA* **98**, 4710–4715 (2001).
36. Parent, J. M., Valentin, V. V. & Lowenstein, D. H. Prolonged seizures increase proliferating neuroblasts in the adult rat subventricular zone–olfactory bulb pathway. *J. Neurosci.* **22**, 3174–3188 (2002).
37. Arvidsson, A., Collin, T., Kirik, D., Kokaia, Z. & Lindvall, O. Neuronal replacement from endogenous precursors in the adult brain after stroke. *Nature Med.* **8**, 963–970 (2002).
38. Nakatomi, H. *et al.* Regeneration of hippocampal pyramidal neurons after ischemic brain injury by recruitment of endogenous neural progenitors. *Cell* **110**, 429–441 (2002).
39. Parent, J. M., Vexler, Z. S., Gong, C., Derugin, N. & Ferriero, D. M. Rat forebrain neurogenesis and striatal neuron replacement after focal stroke. *Ann. Neurol.* **52**, 802–813 (2002).
40. Jin, K. *et al.* Directed migration of neuronal precursors into the ischemic cerebral cortex and striatum. *Mol. Cell. Neurosci.* **24**, 171–189 (2003).
41. Curtis, M. A. *et al.* Increased cell proliferation and neurogenesis in the adult human Huntington's disease brain. *Proc. Natl Acad. Sci. USA* **100**, 9023–9027 (2003).
42. Jin, K. *et al.* Increased hippocampal neurogenesis in Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **101**, 343–347 (2004).
43. Jin, K. *et al.* Enhanced neurogenesis in Alzheimer's disease transgenic (PDGF-APP^{Sw},Ind) mice. *Proc. Natl Acad. Sci. USA* **101**, 13363–13367 (2004).
44. Jin, K. *et al.* Fibroblast growth factor-2 promotes neurogenesis and neuroprotection and prolongs survival in a transgenic mouse model of Huntington disease. *Proc. Natl Acad. Sci. USA* **102**, 18189–18194 (2005).
45. Fallon, J. *et al.* *In vivo* induction of massive proliferation, directed migration, and differentiation of neural cells in the adult mammalian brain. *Proc. Natl Acad. Sci. USA* **97**, 14686–14691 (2000).
46. Zhao, M. *et al.* Evidence for neurogenesis in the adult mammalian substantia nigra. *Proc. Natl Acad. Sci. USA* **100**, 7925–7930 (2003).
47. Chi, L. *et al.* Motor neuron degeneration promotes neural progenitor cell proliferation, migration and neurogenesis in the spinal cords of ALS mice. *Stem Cells* published online 11 August 2005; doi:10.1634/stemcells.2005-0076 (2005).
48. Jin, K. *et al.* Vascular endothelial growth factor (VEGF) stimulates neurogenesis *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* **99**, 11946–11950 (2002).
49. Schanzer, A. *et al.* Direct stimulation of adult neural stem cells *in vitro* and neurogenesis *in vivo* by vascular endothelial growth factor. *Brain Pathol.* **14**, 237–248 (2004).
50. Mani, N., Khaibullina, A., Krum, J. M. & Rosenstein, J. M. Astrocyte growth effects of vascular endothelial growth factor (VEGF) application to perinatal neocortical explants: receptor mediation and signal transduction pathways. *Exp. Neurol.* **192**, 394–406 (2005).
51. Jin, K. *et al.* Cerebral neurogenesis is induced by intranasal administration of growth factors. *Ann. Neurol.* **53**, 405–409 (2003).
52. Cao, L. *et al.* VEGF links hippocampal activity with neurogenesis, learning and memory. *Nature Genet.* **36**, 827–835 (2004).
53. Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4**, 7–25 (1978).
54. Palmer, T. D., Willhoite, A. R. & Gage, F. H. Vascular niche for adult hippocampal neurogenesis. *J. Comp. Neurol.* **425**, 479–494 (2000).
55. Louissaint, A., Rao, S., Leventhal, C. & Goldman, S. A. Coordinated interaction of neurogenesis and angiogenesis in the adult songbird brain. *Neuron* **34**, 945–960 (2002).
56. Shen, Q. *et al.* Endothelial cells stimulate self-renewal and expand neurogenesis of neural stem cells. *Science* **304**, 1338–1340 (2004).
57. Mukoyama, Y. S., Gerber, H. P., Ferrara, N., Gu, C. & Anderson, D. J. Peripheral nerve-derived VEGF promotes arterial differentiation via neuropilin 1-mediated positive feedback. *Development* **132**, 941–952 (2005).
58. Carmeliet, P. & Tessier-Lavigne, M. Common mechanisms of nerve and blood vessel wiring. *Nature* **436**, 193–200 (2005).
59. Haigh, J. J. *et al.* Cortical and retinal defects caused by dosage-dependent reductions in VEGF-A paracrine signaling. *Dev. Biol.* **262**, 225–241 (2003).
60. Sondell, M., Lundborg, G. & Kanje, M. Vascular endothelial growth factor has neurotrophic activity and stimulates axonal outgrowth, enhancing cell survival and Schwann cell proliferation in the peripheral nervous system. *J. Neurosci.* **19**, 5731–5740 (1999).
61. Silverman, W. F., Krum, J. M., Mani, N. & Rosenstein, J. M. Vascular, glial and neuronal effects of vascular endothelial growth factor in mesencephalic explant cultures. *Neuroscience* **90**, 1529–1541 (1999).
62. Rosenstein, J. M., Mani, N., Khaibullina, A. & Krum, J. M. Neurotrophic effects of vascular endothelial growth factor on organotypic cortical explants and primary cortical neurons. *J. Neurosci.* **23**, 11036–11044 (2003).
63. Khaibullina, A. A., Rosenstein, J. M. & Krum, J. M. Vascular endothelial growth factor promotes neurite maturation in primary CNS neuronal cultures. *Dev. Brain Res.* **148**, 59–68 (2004).
64. Jin, K. L. *et al.* Vascular endothelial growth factor stimulates neurite outgrowth from cerebral cortical neurons via rho kinase signalling. *J. Neurobiol.* (in the press).
65. Jin, K. L., Mao, X. O. & Greenberg, D. A. Vascular endothelial growth factor rescues HN33 neural cells from death induced by serum withdrawal. *J. Mol. Neurosci.* **14**, 197–203 (2000).
66. Jin, K. L., Mao, X. O. & Greenberg, D. A. Vascular endothelial growth factor: direct neuroprotective effect in *in vitro* ischemia. *Proc. Natl Acad. Sci. USA* **97**, 10242–10247 (2000).
67. Jin, K. *et al.* Caspase-3 and the regulation of hypoxic neuronal death by vascular endothelial growth factor. *Neuroscience* **108**, 351–358 (2001).
68. Matsuzaki, H. *et al.* Vascular endothelial growth factor rescues hippocampal neurons from glutamate-induced toxicity: signal transduction cascades. *FASEB J.* **15**, 1218–1220 (2001).
69. Svensson, B. *et al.* Vascular endothelial growth factor protects cultured rat hippocampal neurons against hypoxic injury via an antiexcitotoxic, caspase-independent mechanism. *J. Cereb. Blood Flow Metab.* **22**, 1170–1175 (2002).
70. Kovacs, Z., Ikezaki, K., Samoto, K., Inamura, T. & Fukui, M. VEGF and Flt: expression time kinetics in rat brain infarct. *Stroke* **27**, 1865–1873 (1996).
71. Hayashi, T., Abe, K., Suzuki, H. & Itomaya, Y. Rapid induction of vascular endothelial growth factor gene expression after transient middle cerebral artery occlusion in rats. *Stroke* **28**, 2039–2044 (1997).
72. Lennmyr, F., Ata, K. A., Funa, K., Olsson, Y. & Terent, A. Expression of vascular endothelial growth factor (VEGF) and its receptors (Flt-1 and Flk-1) following permanent and transient occlusion of the middle cerebral artery in the rat. *J. Neuropathol. Exp. Neurol.* **57**, 874–882 (1998).
73. Plate, K. H., Beck, H., Danner, S., Allegrini, P. R. & Wiessner, C. Cell type specific upregulation of vascular endothelial growth factor in an MCA-occlusion model of cerebral infarct. *J. Neuropathol. Exp. Neurol.* **58**, 654–666 (1999).
74. Rosenstein, J. M., Mani, N., Silverman, W. F. & Krum, J. M. Patterns of brain angiogenesis after vascular endothelial growth factor administration *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* **95**, 7086–7091 (1998).
75. Krum, J. M., Mani, N. & Rosenstein, J. M. Angiogenic and astroglial responses to vascular endothelial growth factor administration in adult rat brain. *Neuroscience* **110**, 589–604 (2002).
76. Hayashi, T., Abe, K. & Itoyama, Y. Reduction of ischemic damage by application of vascular endothelial growth factor in rat brain after transient ischemia. *J. Cereb. Blood Flow Metab.* **18**, 887–895 (1998).
77. Bao, W. L., Lu, S. D., Wang, H. & Sun, F. Y. Intraventricular vascular endothelial growth factor antibody increases infarct volume following transient cerebral ischemia. *Chung Kuo Yao Li Hsueh Pao* **20**, 313–318 (1999).
78. Sun, Y. *et al.* VEGF-induced neuroprotection, neurogenesis, and angiogenesis after focal cerebral ischemia. *J. Clin. Invest.* **111**, 1843–1851 (2003).
79. Sun, Y. *et al.* Increased severity of cerebral ischemic injury in vascular endothelial growth factor-B (VegfB)-deficient mice. *J. Cereb. Blood Flow Metab.* **24**, 1146–1152 (2004).
80. Kalaria, R. N. *et al.* Vascular endothelial growth factor in Alzheimer's disease and experimental cerebral ischemia. *Brain Res. Mol. Brain Res.* **62**, 101–105 (1998).
81. Del Bo, R. *et al.* Vascular endothelial growth factor gene variability is associated with increased risk for AD. *Ann. Neurol.* **57**, 373–380 (2005).
82. Schratzberger, P. *et al.* Reversal of experimental diabetic neuropathy by VEGF gene transfer. *J. Clin. Invest.* **107**, 1083–1092 (2001).
83. Schratzberger, P. *et al.* Favorable effect of VEGF gene transfer on ischemic peripheral neuropathy. *Nature Med.* **6**, 405–413 (2000).
84. Chattopadhyay, M. *et al.* HSV-mediated gene transfer of vascular endothelial growth factor to dorsal root ganglia prevents diabetic neuropathy. *Gene Therapy* **12**, published online 21 April 2005; doi:10.1038/sj.gt.3302533 (2005).
85. Fig, J. *Luckiest Man: The Life and Death of Lou Gehrig* (Simon & Schuster, New York, 2005).
86. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59–62 (1993).
87. Hadano, S. *et al.* A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nature Genet.* **29**, 166–173 (2001).
88. Yang, Y. *et al.* The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nature Genet.* **29**, 160–165 (2001).
89. Chen, Y. Z. *et al.* DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.* **74**, 1128–1135 (2004).
90. Nishimura, A. L. *et al.* A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *Am. J. Hum. Genet.* **75**, 822–831 (2004).
91. Oosthuyse, B. *et al.* Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration. *Nature Genet.* **28**, 131–138 (2001).
92. Lambrecht, D. *et al.* VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nature Genet.* **34**, 383–394 (2003).
93. Zheng, C., Nennesmo, I., Fadeel, B. & Henter, J. I. Vascular endothelial growth factor prolongs survival in a transgenic mouse model of ALS. *Ann. Neurol.* **56**, 564–567 (2004).
94. Störkebaum, E. *et al.* Treatment of motoneuron degeneration by intracerebroventricular delivery of VEGF in a rat model of ALS. *Nature Neurosci.* **8**, 85–92 (2005).
95. Azzouz, M. *et al.* VEGF delivery with retrogradely transported lentivector prolongs survival in a mouse ALS model. *Nature* **429**, 413–417 (2004).
96. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**, 77–79 (1991).
97. Sopher, B. L. *et al.* Androgen receptor YAC transgenic mice recapitulate SBMA motor neuronopathy and implicate VEGF164 in the motor neuron degeneration. *Neuron* **41**, 687–699 (2004).
98. Hillen, T. *et al.* Cause of stroke recurrence is multifactorial: patterns, risk factors, and outcomes of stroke recurrence in the South London Stroke Register. *Stroke* **34**, 1457–1463 (2003).
99. Raber, J. *et al.* Irradiation attenuates neurogenesis and exacerbates ischemia-induced deficits. *Ann. Neurol.* **55**, 381–389 (2004).

Acknowledgements This work was supported by NIH grants NS44921 (D.A.G.) and AG21980 (K.J.) and by the Buck Institute for Age Research.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to D.A.G. (dgreenberg@buckinstitute.org).

Retinal angiogenesis in development and disease

Ray F. Gariano¹ & Thomas W. Gardner²

The retina has long been regarded as 'an approachable part of the brain' for investigating neurosensory processes. Cell biologists are now capitalizing on the accessibility of the retina to investigate important aspects of developmental angiogenesis, including how it relates to neuronal and glial development, morphogenesis, oxygen sensing and progenitor cells. Pathological angiogenesis also occurs in the retina and is a major feature of leading blinding diseases, particularly diabetic retinopathy. The retina and its clinical disorders have a pivotal role in angiogenesis research and provide model systems in which to investigate neurovascular relationships and angiogenic diseases.

Retinal vascular anatomy, which is highly organized and easily visualized, is central to the attractiveness of the retina as an experimental system. Relationships between retinal vascular and neural structures are apparent in the shared radial orientation of blood vessels and ganglion cell axons, and in planar capillary plexuses that align precisely with horizontal neural and astrocytic laminae (Fig. 1). In addition, vascularized and avascular compartments are strictly segregated in the retina; this feature is strikingly depicted in the human central retina, or fovea, which is entirely devoid of vessels (Fig. 1f).

By contrast, pathological retinal angiogenesis — a key component of irreversible causes of blindness — generates chaotically orientated and physiologically deficient vessels that do not conform to neuronal histology, which can lead to vision-threatening exudation and haemorrhage (Fig. 2). The retina therefore provides a model neurovascular system, attracting the interest of both developmental biologists and clinicians. In recent years, substantial progress has been made towards the twin goals of retinal vascular research: determining mechanisms that induce and pattern the retinal vascular system, and dissecting disease processes that lead to its disintegration. In this review, we highlight advances in these complementary realms.

Retinal vascularization

Retinal vascularization begins in the most superficial (or inner) retinal layers at the optic nerve head, and radiates outwards from this central point. It reaches the retinal periphery just before birth in humans, and during the first week of life in mice^{1,2} (Fig. 3a–c). Additional capillary networks in deeper retinal layers then arise by sprouting from the nascent inner vascular layer (Fig. 3d). As in the brain, angiogenic sprouting is the predominant mechanism of retinal vascularization (reviewed in ref. 3), although additional modes of vascular growth, such as intussusception, are not excluded. Vasculogenesis, in which vessels form by concatenation of vascular precursor cells into solid cords that then lumenize, might contribute to growth of the superficial plexus, but definitive evidence for this is lacking^{3–5}.

More than 50 years ago, Michaelson pioneered the use of dye-perfusion techniques to reveal embryonic and perinatal retinal vasculature. He discovered that capillaries grow profusely when adjacent to

nascent venules, and more sparsely around arteries¹. From these simple observations, early investigators posited that an oxygen-sensitive molecule (termed 'Factor X') controls retinal vascular development — an idea based on then recent concepts of morphogen gradients^{1,6,7}.

Three lines of evidence bolstered the idea that oxygen regulates blood vessel growth in the retina. First, pathologic retinal angiogenesis occurs in several diseases that are characterized by retinal ischaemia. Second, the timing and central-to-peripheral direction of retinal vascularization coincide with developmental processes that presumably determine local oxygen tension. In particular, during late embryogenesis, centrifugal waves of differentiation and proliferation, and subsequently of synaptic connectivity and electrical activity, begin in the central retina and spread peripherally. Extension of the superficial vascular plexus follows in the wake of these waves, suggesting that metabolic demand and attendant 'physiological hypoxia' drive vascular growth⁸. Lowering inspired oxygen in neonatal kittens reduces the rate and density of retinal vascularization⁹. Third, the expression of several angiogenic and angioinhibitory factors is oxygen dependent. Vascular endothelial growth factor (VEGF) is a hypoxia-inducible cytokine that is strongly implicated as the elusive Factor X by virtue of its absolute requirement for retinal vascularization, and its expression in spatial and temporal conjunction with developing retinal blood vessels^{10–13} (Fig. 4). Moreover, mice in which only a single VEGF isoform is expressed exhibit distinct retinal vascular phenotypes consistent with the notion that each isoform acts across limited, but variable, spatial ranges to create the vascular pattern¹⁴. In particular, in mice expressing only VEGF120, vascular coverage of the retina and arteriolar differentiation are impaired, whereas mice expressing VEGF188 have normal veins but aborted arteriolar specification (VEGF164/164 retinas appear normal). It is thought that guidance cues from more cell-bound and matrix-bound VEGF164 and VEGF188 provide a molecular track along which growing retinal vessels appropriately configure. The more highly diffusible VEGF120 fails to provide such spatial information, resulting in reduced vessel branching¹⁴.

Once established, retinal capillaries adjacent to newly formed arterioles are selectively pruned (Fig. 3e, f). This remodelling correlates with reduced VEGF messenger RNA (mRNA) production and endothelial cell apoptosis in these areas. Presumably, relative hyperoxia surrounding arter-

¹Department of Ophthalmology, A-157, 300 Pasteur Drive, Stanford University School of Medicine, Palo Alto, California 94305, USA; ²Departments of Ophthalmology and Cellular & Molecular Physiology, Penn State College of Medicine, Hershey, Pennsylvania 17033, USA.

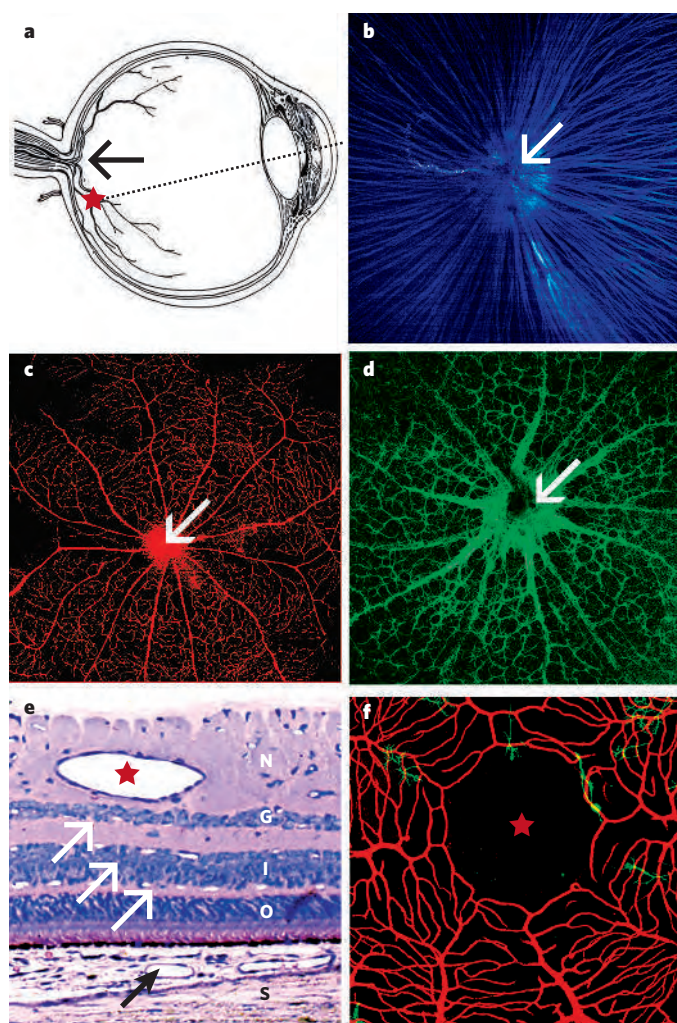


Figure 1 | Anatomy and histological organization of the retina. **a**, Schematic cross-section through the human eye showing the vascularized retina lining the inner surface of the posterior two-thirds of the eye. The optic nerve (arrow) exits the eye at a slightly eccentric position. In primates, the central visual axis (dotted line) is centred on the fovea (star; see **f**). **b–d**, Similar distributions of blood vessels, neurons and glia in a whole-mounted mouse retina. **b**, Radially orientated ganglion cell axons (labelled blue for leptin receptor) exit the eye through the optic nerve (arrow). **c**, Fluorescent dextran (red) angiogram of adult retina; blood vessels also radiate from the optic nerve (white arrow) to the periphery. **d**, Retinal astrocyte meshwork labelled for glial fibrillary acidic protein (green) resembling that of blood vessels. **e**, Human retinal cross-section showing capillaries and larger vessels (star) in the superficial nerve fibre layer (N). Deeper capillary networks (arrows) align within the ganglion cell (G) layer and along each side of the inner nuclear layer (I); the outer nuclear layer (O) is avascular, and receives blood from choroidal vessels (black arrow) between the retina and the sclera (S), which is the outer white coat of the eye. Pathological angiogenesis is not confined to specific neuronal layers, and can grow into the vitreous cavity or outer retina. **f**, In humans, a circular avascular zone of about 450 μm at the fovea (see star in **a**) improves central vision by reducing light scatter from blood vessels.

ies negatively regulates VEGF production, and reduces its actions as an endothelial cell survival factor^{15,16}. Immune cells entering the eye from the circulation also participate in remodelling of the nascent retinal vasculature. An outward wave of T cells and other leukocytes adheres to retinal endothelial cells through intercellular adhesion molecule-1 and CD18, and induces loss of capillary segments by Fas ligand-mediated endothelial cell apoptosis¹⁷. Interestingly, remodelling occurs in perinatal retinal explants, so additional mechanisms of vessel retraction must exist that are independent of systemically supplied oxygen or immune cells¹⁸.

VEGF and its receptors are expressed in a non-overlapping manner during retinal vascularization, providing additional mechanisms for vascular patterning. Although VEGF itself is widely expressed in the developing retina, its various receptors (FLK1, FLT1, and neuropilins 1 and 2) appear in a temporally and spatially distinct fashion^{19,20} (Fig. 4a). These receptors mediate discrete functions, such as proliferation, migration, guidance, survival and permeability, so that within areas of VEGF expression in the developing retina, its actions might be delimited by the emergence of receptors, as well as by ligand availability and subtype. Moreover, signals initiated by the same receptor can elicit distinct effects in endothelial cells located at different sites in the growing retinal vasculature. For example, FLK or VEGF receptor 2 (VEGFR-2) activation elicits migration and lamellipodial protrusion in delta-like-4-positive endothelial tip cells at the leading edge of retinal vascularization, whereas it stimulates proliferation in the more proximal stalk cells²¹ (Fig. 4b–d).

Therefore, a single growth factor sculpts the developing retinal vasculature in several ways, including spatiotemporal changes in VEGF and VEGFR distribution, differential effects of VEGF isoforms, distinct spatial range of action and guidance functions of VEGF isoforms, and alternate transduction of VEGFR activation on different endothelial cells. The interplay between VEGF and other angiogenic and angi-inhibitory factors provides further opportunity to generate complex vascular networks^{22–25}.

A key question is how growing vessels sense and respond to varying tissue oxygen levels. A partial answer comes from an inherited disorder, von Hippel–Lindau (VHL) disease, in which vascular endothelial tumours, or haemangiomas, arise in the retina and other organs (Fig. 2b). The VHL protein (pVHL) is a tumour suppressor and functions within an oxygen-sensing pathway. Oxygen-dependent Egl9 enzymes hydroxylate a proline residue on hypoxia-inducible factor 1 α (HIF1 α), which promotes binding of HIF1 α with pVHL. An elongin–BC complex then polyubiquitinates HIF1 α , targeting it to the proteasome for destruction^{26,27}. Under hypoxic conditions, prolyl hydroxylation fails and HIF1 α is spared. It then dimerizes with HIF1 β , translocates to the nucleus and binds to hypoxia response elements within the promoters of several genes. The genetic response to HIF coordinates metabolic and vascular events — including angiogenesis — that generally improve fitness in the face of low oxygen concentrations. In patients and animals with VHL disease, pVHL mutations might dissociate HIF-induced genetic responses from tissue oxygen, allowing angiogenesis unchecked by oxygen²⁸. VHL dysfunction appears to be an important step in the growth of angiogenic tumours.

A role for HIF–VHL in retinal vascularization is suggested by *Vhl* expression within the developing retina²⁹ and by retinal vascular abnormalities in *Hif2 α* -knockout mice³⁰. Experimental retinal angiogenesis correlates with HIF1 α levels and is inhibited by adenoviral delivery of the *Vhl* gene³¹.

Cellular relationships in vascular guidance and patterning

Despite indirect evidence implicating hypoxia as a stimulus for retinal vascularization, oxygen profiles cannot explain the subtle meshwork pattern of retinal blood vessels, and their partial alignment with ganglion cell axons, astrocytes and neuronal elements (Fig. 1). How do these similar cellular distributions arise?

Astrocyte precursors enter the retina through the optic nerve and radiate towards the periphery^{32,33}. Vascularization subsequently follows along the pre-existing astrocytic meshwork (Fig. 5a), suggesting a template function for glial fibres^{32–35}. Consistent with this, fine endothelial filipodia at the tips of growing retinal vessels often extend along processes of underlying astrocytes that secrete VEGF (Fig. 5b, c). Interactions between endothelial and glial processes depend on the cell surface-adhesion molecule R-cadherin. Disruption of homomeric or heteromeric interactions between endothelial and glial cadherins with function-blocking antisera results in a stunted retinal vasculature, which fails to recapitulate the astrocytic pattern and migrates into the normally avascular outer retina³⁴.

If astrocytes serve as a template for growing vessels, what guides the

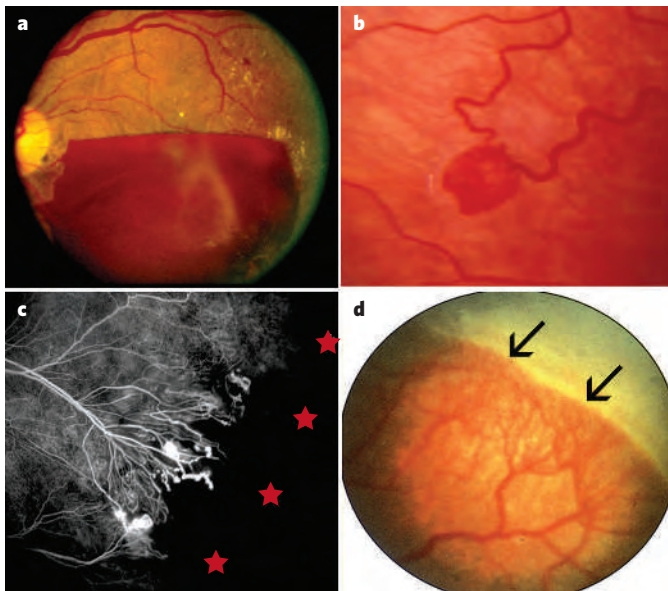


Figure 2 | Clinical retinal diseases characterized by aberrant retinal angiogenesis. **a**, Pre-retinal haemorrhage arising from neovascularization in a patient with diabetic retinopathy obscures a portion of the posterior retina; yellow spots are lipid exudates from hyper-permeable retinal vessels (optic nerve on the left). **b**, An angioma (arrow) connected to the retinal vasculature by feeder and drainer vessels in a patient with von Hippel-Lindau syndrome. **c**, Fluorescein angiogram showing telangiectatic neovascularization at the border of the vascularized posterior (left), and non-perfused peripheral retina (right; stars) in a patient with familial exudative vitreoretinopathy. **d**, Developmental vascularization stops short of the periphery in an infant with retinopathy of prematurity; angiogenesis arises just posterior to the avascular region (arrows).

astrocytes? One factor might be the pre-existing radially orientated axon bundles of ganglion cells, which are the retinal output neurons that project through the optic nerve to visual centres of the brain (Fig. 1a, b). Astrocyte growth roughly follows this radial plan^{32–34} in a process that requires platelet-derived growth factor (PDGF) synthesized by retinal ganglion cells³⁶. Another clue is provided by TLX, which is an orphan receptor with established roles in neural development. Studies of *Tlx*-knockout mice show that it is necessary for normal growth and differentiation of retinal glia; astrocytes are less dense and more coarsely distributed in *Tlx*^{−/−} retinas than in wild-type retinas, and fail to normally express VEGF and R-cadherin³⁷. Notably, the astrocytic

plexus across which vessels grow is also disrupted in VEGF120 mice, indicating that VEGF participates in astrocytic patterning¹⁴.

Congruent cellular patterns in the retina might therefore result from sequential reciprocal cellular interactions: neuron-derived PDGF stimulates and patterns astrocyte invasion, which in turn induces VEGF-dependent and cadherin-dependent vascular growth and guidance. Ingrowth of vessels, perhaps by raising tissue oxygen tension, then promotes astrocytic maturation, as indicated by changes in glial morphology and gene expression. Reduction of VEGF synthesis by astrocytes follows (Fig. 5d, e), and provides a local feedback mechanism to limit further vascular growth³⁸.

Pericytes are smooth-muscle cells that are intimately associated with developing blood vessels; in the retina, they are recruited and differentiated from an unknown precursor cell pool in a PDGFB-dependent manner³⁹. Pericytes associated with the advancing front of developing retinal vessels promote vascular growth by secreting VEGF⁴⁰, and by inducing VEGFR-1 in endothelial cells through transforming growth factor- β 1 (TGF- β 1)⁴¹. Pericyte-endothelial contact might then stabilize vessels by activation of latent TGF- β (ref. 42). Loss of pericytes occurs in diabetic retinopathy and other ischaemic retinal conditions, and diabetic-like retinopathy develops in mice with reduced PDGF activity and pericyte coverage of retinal blood vessels^{41,43}. Pericyte dysfunction might therefore allow a range of vascular changes, including loss of normally growing vessels (by reduction of VEGF and VEGFR), loss of formed vessels (by reduction of VEGF endothelial survival actions) and abnormal angiogenesis (by removal of the inhibitory effect of TGF- α on endothelial proliferation). Microglia, which are phagocytic dendritic cells with immune functions, co-distribute with retinal blood vessels during development⁴⁴ and are activated in ischaemic diseases. However, their role in retinal vascular development, if any, is unknown.

In contrast to the finely graded cellular interactions during development, astrocytic responsiveness to experimentally induced hypoxia is exaggerated, and leads to degeneration and focal disruptions of the glial ensheathment surrounding retinal blood vessels⁴⁵. Neovascularization develops at these sites and can extend beyond the retinal surface into the vitreous cavity, indicating that astrocytes might stabilize already formed vessels and restrict them to appropriate retinal layers. Therefore, whereas both astrocytes and pericytes promote normal vascularization, glial degeneration and pericyte loss are permissive conditions for pathological angiogenesis in ischaemic retina⁴⁵.

Blood vessels and glia often use attractive and repulsive guidance cues employed by axons²². Similar configurations of retinal neurons, astrocytes and endothelial cells might result from consecutive operation of these processes within a shared microenvironment. If so, the abnormal location and orientation of pathological retinal neovessels

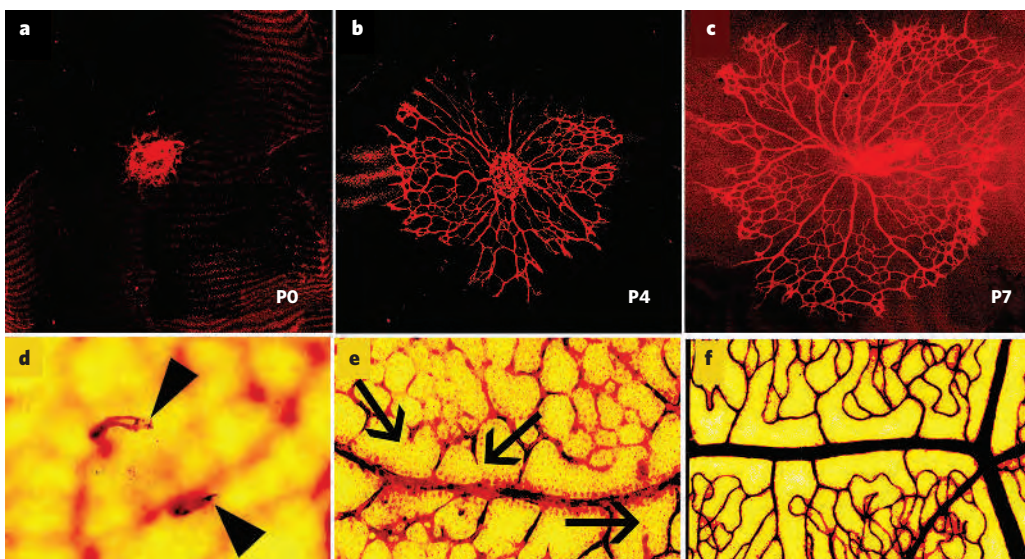


Figure 3 | Stereotypical timing and morphology of retinal vascularization. **a–c** In mice, retinal vessels (red) arise from the optic nerve around birth (P0) then extend radially in the superficial retina over 7–10 days to reach the periphery. **d**, In primates, deeper capillary networks form by endothelial sprouting (arrowheads) from the previously formed superficial vascular network (blurred, in background). **e**, In primates, shortly after retinal vessels form, capillary segments adjacent to nascent arteries (arrows) retract, to yield a periarterial capillary-free zone.

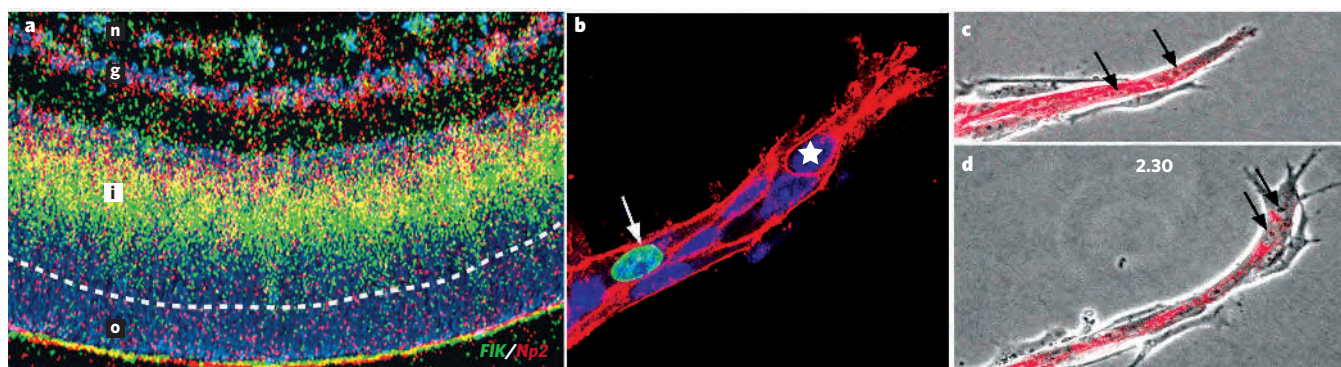


Figure 4 | Vascular endothelial growth factor in retinal development. **a**, Non-overlapping distribution of messenger RNA for VEGFRs FLK (green) and neuropilin-2 (red) in perinatal mouse retina; the dotted line separates the inner (i) and outer (o) nuclear layers. g, ganglion cell layer; n, nerve fibre layer; cells are stained blue. **b**, In cultured vascular structures exposed to VEGF, proliferation occurs in endothelial stalk cells (labelled green for phospho-histone-H3) but not in distal tip cells (star). **c, d**, The endothelial tip cell (arrows) within a single sprout labelled red for CD31 migrates and extends filopodial processes in response to VEGF at baseline (**c**) and 2.5 min after exposure (**d**).

partly represents the application of guidance mechanisms to an already altered (diseased) environment. This perspective suggests that in conditions such as diabetic retinopathy, neuronal, glial and pericyte damage might precede ischaemia and angiogenesis, and reflect an underlying pathological process with broad cellular effects (see below).

An alternative explanation for similarly patterned retinal glial and vascular networks is that both arise in response to mechanical forces within a shared structural milieu. Mechanical stress generated by cytoskeletal activity in cultured endothelial cells transmits to underlying matrigel via cell-matrix attachments, resulting in curvilinear deformations in the matrix⁴⁶. These tension lines seem to function as 'tracks' along which endothelial cells then migrate. The meshwork pattern of matrix-tension tracks *in vitro* is similar to that of growing retinal vessels and their filopodia. Investigators have described cystic spaces within the superficial retina, through which vessels and astrocytes migrate during development³⁻⁶. It is conceivable that endothelial cells and astrocytes behave in the retina as they do *ex vivo*, generating vascular and glial networks along tension tracks formed as they migrate within, and attach to, this hypocellular and protein-rich compartment.

Novel retinal disease mutations and vascular development

Further clues to retinal vascular organization, particularly confinement of blood vessels to inner retinal layers, come from investigations of rare diseases. Norrie's disease is characterized by retinal dysplasia and abnormal retinal vascularization, and mice with targeted disruption of the gene that encodes norrin protein lack deep intraretinal vessels⁴⁷.

Interestingly, mice deficient in *frizzled-4* have retinal vascular defects similar to those seen in norrin mutants, and *frizzled-4* mutations occur in patients with familial exudative vitreoretinopathy⁴⁷⁻⁴⁹, which is an autosomal disorder characterized by incomplete peripheral retinal vascularization (Fig. 2c). In addition, defective lipoprotein receptor-related protein 5 is associated with similar retinal vascular defects in mice⁴⁹.

Recent studies unify these findings within the Wnt signalling system. Norrin acts as a secreted activator of canonical Wnt- β -catenin signalling by binding to frizzled-4 (itself a Wnt receptor), possibly in conjunction with lipoprotein receptor-related proteins (Wnt co-receptors)^{48,50}. A description of norrin and frizzled-4 expression in relation to developing retinal blood vessels might clarify why only peripheral vasculature is lost in patients with familial exudative vitreoretinopathy, or why the deeper retinal capillaries are specifically aborted in mice with Wnt-related mutations. Wnt-mediated processes might also relate to observations suggesting that, in humans, vasculogenic growth occurs in the central superficial retina, whereas angiogenic sprouting vascularizes the peripheral and deeper retina (see above)^{3-5,51}. Norrin-defective and frizzled-4-defective mice also exhibit cochlear vascular defects⁴⁷, reminiscent of sensorineural hearing loss in Norrie's disease, whereas mice with mutations in frizzled-5 and lipid phosphatase LPP3 (another regulator of the Wnt/ β -catenin pathway) exhibit defective placental vasculogenesis. Specific components of Wnt signalling might therefore function at different sites of vascular development (see ref. 50 for a discussion).

Ectopic norrin restores retinal vessels in norrin-mutant mice⁵², raising the possibility of a new anti-angiogenic pharmacology for

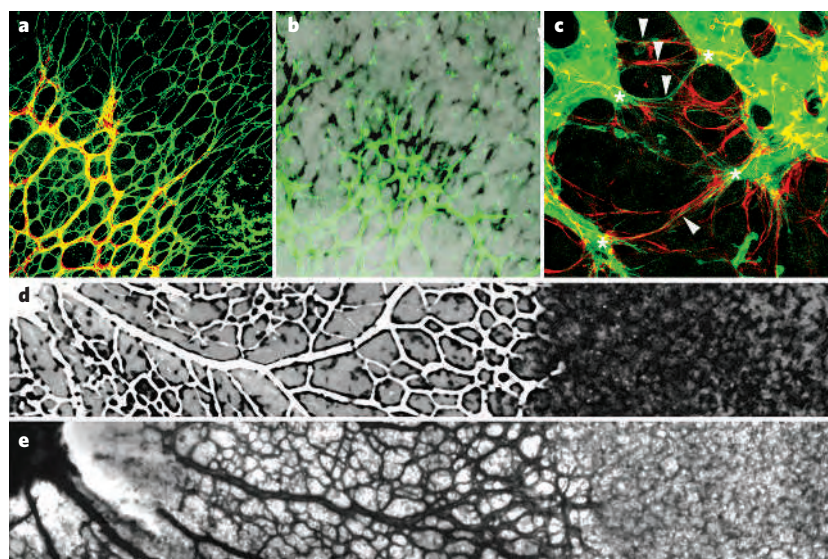


Figure 5 | Glial-vascular relationships in retinal vascularization. **a**, Retinal blood vessels (red, growing towards the upper right) develop along a pre-existing astrocytic meshwork (green, double-label yellow-orange). **b**, VEGF expression (black) by underlying astrocytes is greatest just in advance of the most distal growing blood vessels (green). **c**, Filopodial processes from growing retinal vessels (green) are in register with astrocytic processes (red). **d**, VEGF expression (black) is higher in astrocytes beyond the extent of retinal vascularization (right) than in astrocytes where vessels (white) are present. **e**, Astrocytes express glial fibrillary acidic protein (black) more intensely in vascularized (left) than avascular (far right) retinas.

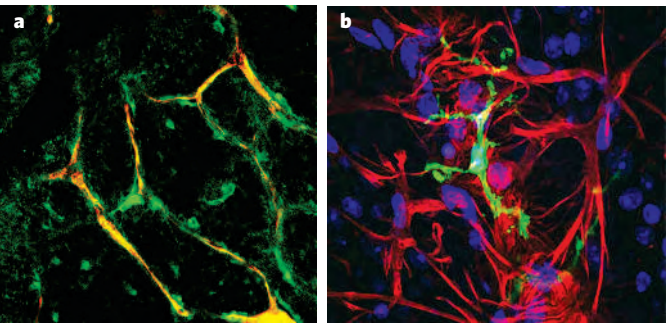


Figure 6 | Bone-marrow-derived progenitor cells incorporate into perinatal and adult retinal vasculature. **a**, Both native (red) endothelial cells and intraocularly injected marrow-derived cells (green, double-label orange-yellow) contribute to newly formed retinal blood vessels in perinatal mice. **b**, Donor marrow-derived cells in adult retina appear as glial or microglial cells (green) overlaying a blood vessel and adjacent astrocytes (red).

retinopathy of prematurity, which is a major cause of infantile blindness. In this disease of prematurely born infants receiving oxygen therapy, vascularization fails to reach the retinal periphery (Fig. 2d). The non-perfused peripheral retina becomes hypoxic and neovascularization can ensue, with complications of haemorrhage, fibrosis and loss of vision. A commonly employed retinal angiogenesis model in neonatal mice shares features with human retinopathy of prematurity⁵³, and might be useful to test whether Wnt signalling factors can restore peripheral retinal vascularization.

A different retinal vascular phenotype occurs in mice with a defect in the very-low-density lipoprotein receptor (VLDLR). In these mice, superficial and deeper retinal blood vessels form as expected, but then ectopic sprouting of small angiomatic growths invades the normally avascular outer retina⁵⁴. VLDLR interacts with reelin and apolipoprotein E to guide neurons to their proper cortical laminae during brain development⁵⁵, and it is tempting to speculate that it has a corresponding guidance function for retinal vessels⁵⁶. Guidance strategies among neural and endothelial cells have been recently reviewed²². Intraretinal neovascularization in VLDLR-mutant mice resembles that seen in a subset of patients with age-related macular degeneration, in which angiomatic branching from inner retinal vessels extends into the outer retina and subretinal space.

Progenitor cells in retinal vascularization

A population of lineage-negative cells in the bone marrow mobilizes in response to various stimuli and incorporates into growing blood vessels at sites of ischaemic or other damage. When injected systemically into adult mice, these cells incorporate into experimentally

induced retinal and choroidal neovascularization^{57,58}. Similarly, when injected into the eyes of mice shortly after birth, lineage-negative bone marrow cells incorporate into developing retinal vessels, resulting in a mosaic vasculature comprising native and exogenous cells⁵⁹ (Fig. 6a). Interestingly, these cells can restore lost capillaries and reduce neuronal loss in a mouse model of retinal vascular and neuronal degeneration⁶⁰. Interactions between R-cadherin on marrow-derived precursor cells and retinal cell-surface cadherins are required to target these cells to developing retinal blood vessels⁶¹, reminiscent of R-cadherin-mediated linkage of retinal vessels to an astrocytic template during normal development³⁴.

At present, it is unclear whether haematopoietic stem cells or other types of progenitor cell contribute normally to retinal vascular development. Certainly, an ongoing supply of circulating cells seems unnecessary as vascularization occurs in retinal explants disconnected from the systemic circulation shortly after birth¹⁸. The ability to visualize site-specific and stage-specific expression of stem cell markers might resolve this issue. Marrow-derived cells have been shown to associate with adult retinal blood vessels after sublethal radiation and marrow reconstitution, although probably as perivascular glia rather than as endothelial cells⁶² (Fig. 6b); corresponding studies assessing homing of marrow-derived cells during perinatal retinal vascularization have not been reported.

Diabetic retinopathy

Diabetic retinopathy remains the most common cause of vision impairment in working-age adults in the United States and Europe, and retinal neovascularization occurs in up to 20% of patients with diabetes⁶³. The current treatment was first introduced in the 1950s, but the fundamental approach of ablating the peripheral retina has changed little over the past five decades. This surgical therapy, while effective in reducing the risk of severe vision loss, is only applied after the onset of neovascularization, does not address the basic biological abnormality that leads to this complication, reduces peripheral and night vision, and is uncomfortable and expensive. It is the ocular equivalent of lower-extremity amputation and, so far, there is no clinically proven non-surgical alternative.

Angiogenic factors in proliferative diabetic retinopathy

Pathological retinal neovascularization in patients with diabetes results from an imbalance of pro-angiogenic and anti-angiogenic factors. Since the seminal discovery of VEGF accumulation in eyes with diabetic retinal neovascularization^{64,65}, changes in numerous other cytokines, chemokines, adhesion molecules, vasoactive hormones and immune cells have been reported (Table 1). Together, these changes constitute a complex inflammatory process that results in an aberrant wound-healing response. Most reports emphasize pro-angiogenic stimuli, whereas the natural inhibitory elements have received little

Table 1 | Vitreous and serum factors altered in human proliferative diabetic retinopathy

Increased in vitreous and/or retina	Pro-angiogenic Peptide growth factors: VEGF, HGF, FGF5, leptin, IGF1, IGF2, PDGFAB, SDF1, angiogenin Extracellular matrix-adhesion molecules: ICAM1, oncofetal fibronectin Inflammatory cytokines: IL-6, IL-8, ET-1, TNF-α, TGF-β1, AGEs Complement: complement C(4) fragment Polyamines: spermine, spermidine Vasoactive peptides: endothelin-1, angiotensin-2, angiotensin-2, adrenomedullin, ACE, nitrate Inflammatory cells: CD4 and CD8 (T lymphocytes), CD22 (B lymphocytes), macrophages, HLA-DR
Increased in vitreous and/or retina	Anti-angiogenic Endostatin, angiostatin, PEDF, TGF-β1 Undefined retinal function: α1-antitrypsin, α2-HS glycoprotein
Decreased in vitreous and/or retina	Angiotensin-2, putrescine, kallistatin, chymase, TGF-β2 activation, CD55, CD59
No change in vitreous and/or retina	ACE, C1q and C4
Increased in serum	NO, sIL-2R, IL-8, TNF-α, VEGF, angiotensin-2, renin, endothelin
Decreased in serum	Soluble angiotensin receptor Tie2, IL-1β, IL-6

References available from authors on request. ACE, angiotensin-converting enzyme; AGE, advanced glycation end-products; FGF, fibroblast growth factor; HGF, hepatocyte growth factor (scatter factor); HLA, human leukocyte antigen; ICAM, intercellular adhesion molecule; IGF, insulin-like growth factor; IL, interleukin; NO, nitric oxide; PDGF, platelet-derived growth factor; PEDF, pigment epithelium-derived factor; SDF1, stromal-derived factor-1; sIL-2R, soluble interleukin-2 receptor; TGF-α1, transforming growth factor-α1; TNF-β, tumour necrosis factor-β; VEGF, vascular endothelial growth factor.

attention. A proteomic study of vitreous fluid from patients with diabetic retinopathy detected 56 proteins: 29 of these were vitreous-specific and the balance were plasma derived⁶⁶. TNF- α , endothelin and insulin-like growth factors are increased in plasma, and accumulate in the vitreous cavity as a result of blood-retinal barrier breakdown; therefore, systemic inflammation clearly contributes to the ocular disease. At this point, it is impossible to know which of these factors is causative, because temporal changes are unknown. Moreover, lipid or carbohydrate mediators that probably play a role⁶⁷ have received little attention. The vitreous levels of growth factors are instructive, but full understanding of the biological activities of receptors, signalling pathways and interactions is needed to achieve a complete understanding of neovascular retinopathy. For example, interleukin-1 α (IL-1 α) is increased, but so is the soluble IL-1 receptor that inhibits IL-1 action, so the net effect is uncertain.

The factors listed in Table 1 undoubtedly act at different stages of the inflammatory/anti-inflammatory process, or relate more to fibrotic processes that occur in diabetic eyes. In addition to direct effects on vascular cells, inflammatory mediators probably also modify the structure of the vitreous gel that fills the bulk of the eye. Changes in vitreous structure might in turn stimulate vessel growth and fibrosis. Unfortunately, it is difficult to study this problem because the retina cannot be biopsied in humans to provide tissue samples for analysis, and animal models of diabetes do not develop neovascularization; therefore, inferences extrapolated from retinopathy of prematurity and retinal vein-occlusion models might not accurately predict responses in a chronic human disease. Moreover, it remains unclear which aspects of the neovascular process are directly related to diabetes, which are non-specific wound-healing changes, which if any are ocular-specific or systemically derived, and how the process might differ in juvenile versus adult-onset diabetic patients. Clearly, as no single factor is causative or permissive for human diabetic retinopathy, it remains a challenge to determine the combinations of molecules or cells that provide the best therapeutic targets.

Is the retina really hypoxic in diabetes?

Given the fundamental roles of blood vessels in supplying nutrients (oxygen, amino acids, glucose and fatty acids) and removing waste products, which tissue abnormalities might pathological retinal angiogenesis attempt to compensate? Most studies have emphasized the role of hypoxia, because clinical retinal angiography shows non-perfused capillaries (Fig. 2); however, the direct empirical evidence for reduced retinal oxygen tension in human diabetic retinopathy is surprisingly limited. Retinal oxygen saturation decreases in patients with diabetes and no retinopathy, implying a mismatch of supply and consumption⁶⁸. At sites of chorioretinal scars from laser treatment, retinal oxygen tension is elevated compared with untreated regions⁶⁹. Hyperoxia improves contrast sensitivity⁷⁰ and might improve certain features of diabetic retinopathy⁷¹, but pre-retinal oxygenation is equivalent in cats and dogs with less than 1 year of diabetes versus controls^{72,73}. One study found reduced retinal oxygen partial pressure in cats with 6–8 years of diabetes⁷⁴. No studies have directly demonstrated reduction of retinal oxygen levels in humans with diabetes compared with controls (E. Stefansson, personal communication). Therefore, while substantial indirect evidence argues for retinal hypoxia, current data do not establish a causal relationship between retinal hypoxia due to capillary occlusion and retinal neovascularization in diabetes.

The evidence for retinal tissue hypoxia in animal models is based largely on the findings of overexpression of peptide growth factors that are regulated by HIF, and retinal HIF activity is increased in diabetic rats⁷⁵. However, VEGF, HIF, PDGF and other peptide growth factors are also part of a common response to many cellular stresses. Recent work shows that all retinal cell types, including neurons, glial, microglial and vascular cells, are affected by diabetes, resulting in a neurovascular disorder⁷⁶. The retinal neurovascular degeneration of diabetes includes neuronal and vascular cell apoptosis⁷⁷, and microglial and glial cell activation, which provides intraretinal sources of cytokines and chemokines^{78,79}. If the teleologic goal of the organism is

to maintain neuron-dependent vision in the presence of injury, then the neovascular lesions could partly compensate for metabolic derangements in neural retina. That is, VEGF could increase initially to provide trophic support to neurons through neuronal VEGFRs^{19,20,22} (Fig. 4a), but at the cost of increased vascular permeability. If the stress persists, as in diabetes, the compensatory response might predominate and contribute to the clinical picture of diabetic retinopathy, including neovascularization and oedema due to vascular leakage.

The endoplasmic reticulum (ER) stress response is a potent reaction to a variety of cellular injuries, including viral infection, growth factor and nutrient deprivation, and hypoxia, which are associated with cell death and vascular responses. ER stress responses occur as a result of nutrient deprivation, excess lipid accumulation and glucose deficiency, and coordinate the balance between cell survival and death signals; these include regulation of VEGF and PDGF expression at the level of mRNA translation⁸⁰. Therefore, regardless of whether it is due to true tissue hypoxia or, perhaps, neural retinal sensing of cellular 'malnourishment' owing to various metabolic aspects of diabetes, ER stress might be a reasonable candidate to help explain certain features of neovascular diabetic retinopathy.

Perspectives

Several caveats should be considered in developing treatments for retinopathy. First, angiogenesis in isolation or in normal development is different from that in the context of multiple systemic and local disease-related derangements, or in humans with diverse genetic and cultural backgrounds; pathological retinal angiogenesis does not simply recapitulate developmental mechanisms. Second, anti-VEGF therapies, which are already used to treat ocular neovascularization in age-related macular degeneration, raise concerns for diabetic patients; antagonism of VEGF might interfere with myocardial revascularization in patients who are already at high risk for cardiac ischaemia, for example, and loss of the neurotrophic²² and vasculotrophic¹⁶ actions of VEGF might exacerbate neuronal loss and ischaemia in diabetic eyes. Third, therapies that address the multifactorial nature of retinopathy will probably be more successful than single-molecule-specific approaches.

Not surprisingly, at this point, more questions than answers remain, but the field is poised to move forward rapidly. The diabetes epidemic demands more effective and less expensive biologically based treatments for proliferative diabetic retinopathy. Achieving this goal will require more comprehensive analyses of retinal and vitreous composition, greater understanding of retinal nutrition by oxygen and other substrates, better animal models, and interactions between experts in angiogenesis, immunology, wound-healing and clinical retinal disorders.

The study of diabetic retinopathy and other retinal disorders opens new lines of angiogenesis inquiry, and developmental retinal angiogenesis models are crucial for investigating neurovascular relationships and bringing anti-angiogenic therapies to patients. These complementary approaches have led us from initial theories emphasizing oxygen as a primary determinant of retinal angiogenesis to a large and growing field that explores metabolic, immune, glial, neuronal, gene expression, and perivascular and progenitor cell factors and gives hope to those who suffer from blinding angiogenic diseases. ■

1. Michaelson, I. C. The mode of development of the vascular system of the retina, with some observations on its significance for certain retinal diseases. *Trans. Ophthalmol. Soc. UK* **68**, 137–181 (1948).
2. Connolly, S. E., Hores, T. A., Smith, L. E. & D'Amore, P. A. Characterization of vascular development in the mouse retina. *Microvasc. Res.* **36**, 275–290 (1988).
3. Gariano, R. F. Cellular mechanisms in retinal vascular development. *Prog. Retinal Eye Res.* **22**, 295–306 (2003).
4. Flower, R. W., McLeod, D. S., Luty, G. A., Goldberg, B. & Wajner, S. D. Postnatal retinal vascular development of the puppy. *Invest. Ophthalmol. Vis. Sci.* **26**, 957–968 (1985).
5. Chan-Ling, T. L., Halasz, P. & Stone, J. Development of retinal vasculature in the cat: processes and mechanisms. *Curr. Eye Res.* **9**, 459–478 (1990).
6. Ashton, N. Oxygen and the growth and development of retinal vessels. *In vivo and in vitro studies. The XX Francis I. Proctor Lecture. Am. J. Ophthalmol.* **62**, 412–435 (1966).
7. Wise, G. N. Factors influencing retinal new vessel formation. *Am. J. Ophthalmol.* **52**, 637–650 (1961).
8. Chan-Ling, T., Gock, B. & Stone, J. The effect of oxygen on vasoformative cell division. Evidence that 'physiological hypoxia' is the stimulus for normal retinal vasculogenesis. *Invest. Ophthalmol. Vis. Sci.* **36**, 1201–1214 (1995).

9. Phelps, D. L. Oxygen and developmental retinal capillary remodeling in the kitten. *Invest. Ophthalmol. Vis. Sci.* **31**, 2194–2200 (1990).
10. Stone, J. et al. Development of retinal vasculature is mediated by hypoxia-induced vascular endothelial growth factor (VEGF) expression by neuroglia. *J. Neurosci.* **15**, 4738–4747 (1995).
11. Ozaki, H. et al. Blockade of vascular endothelial cell growth factor receptor signaling is sufficient to completely prevent retinal neovascularization. *Am. J. Pathol.* **156**, 697–707 (2000).
12. Provis, J. M. et al. Development of the human retinal vasculature: cellular relations and VEGF expression. *Exp. Eye Res.* **65**, 555–568 (1997).
13. Pierce, E. A., Avery, R. L., Foley, E. D., Aiello, L. P. & Smith, L. E. Vascular endothelial growth factor/vascular permeability factor expression in a mouse model of retinal neovascularization. *Proc. Natl Acad. Sci. USA* **92**, 905–909 (1995).
14. Stalmans, I. et al. Arteriolar and venular patterning in retinas of mice selectively expressing VEGF isoforms. *J. Clin. Invest.* **109**, 327–336 (2002).
15. Claxton, S. & Fruttiger, M. Role of arteries in oxygen induced vaso-obliteration. *Exp. Eye Res.* **77**, 305–311 (2003).
16. Alon, T. et al. Vascular endothelial growth factor acts as a survival factor for newly formed retinal vessels and has implications for retinopathy of prematurity. *Nature Med.* **1**, 1024–1028 (1995).
17. Ishida, S. et al. Leukocytes mediate retinal vascular remodeling during development and vaso-obliteration in disease. *Nature Med.* **9**, 781–788 (2003).
18. Curatola, A. M., Moscatelli, D., Norris, A. & Hendricks-Munoz, K. Retinal blood vessels develop in response to local VEGF-A signals in the absence of blood flow. *Exp. Eye Res.* **81**, 147–158 (2005).
19. Shih, S. C., Ju, M., Liu, N. & Smith, L. E. Selective stimulation of VEGFR-1 prevents oxygen-induced retinal vascular degeneration in retinopathy of prematurity. *J. Clin. Invest.* **112**, 50–57 (2003).
20. Gariano, R. F., Hu, D. & Helms, J. Expression of angiogenesis-related genes during retinal development. *Mech. Dev. Gene Exp. Patterns* (in the press).
21. Gerhardt, H. et al. VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163–1177 (2003).
22. Carmeliet, P. & Tessier-Lavigne, M. Common mechanisms of nerve and blood vessel wiring. *Nature* **436**, 193–200 (2005).
23. Feeney, S. A. et al. Role of vascular endothelial growth factor and placental growth factors during retinal vascular development and hyaloid regression. *Invest. Ophthalmol. Vis. Sci.* **44**, 839–847 (2003).
24. Sarlos, S. et al. Retinal angiogenesis is mediated by an interaction between the angiotensin type 2 receptor, VEGF, and angiotensin. *Am. J. Pathol.* **163**, 879–887 (2003).
25. Dawson, D. W. et al. Pigment epithelium-derived factor: a potent inhibitor of angiogenesis. *Science* **285**, 245–248 (1999).
26. Jaakkola, P. et al. Targeting of HIF-4 to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* **292**, 468–472 (2001).
27. Ohh, M. et al. Ubiquitination of hypoxia-inducible factor requires direct binding to the M-domain of the von Hippel-Lindau protein. *Nature Cell Biol.* **2**, 423–427 (2000).
28. Rankin, E. B. et al. Inactivation of the arylhydrocarbon receptor nuclear translocator (Arnt) suppresses von Hippel-Lindau disease-associated vascular tumors in mice. *Mol. Cell Biol.* **25**, 3163–3172 (2005).
29. Ma, W. et al. Hepatic vascular tumors, angiectasis in multiple organs, and impaired spermatogenesis in mice with conditional inactivation of the VHL gene. *Cancer Res.* **63**, 5320–5328 (2003).
30. Ding, K., Scortegagna, M., Seaman, R., Birch, D. G. & Garcia, J. A. Retinal disease in mice lacking hypoxia-inducible transcription factor-24. *Invest. Ophthalmol. Vis. Sci.* **46**, 1010–1016 (2005).
31. Ozaki, H. et al. Hypoxia inducible factor-14 is increased in ischemic retina: temporal and spatial correlation with VEGF expression. *Invest. Ophthalmol. Vis. Sci.* **40**, 182–189 (1999).
32. Watanabe, T. & Raff, M. C. Retinal astrocytes are immigrants from the optic nerve. *Nature* **332**, 834–837 (1988).
33. Gariano, R. F., Sage, E. H., Kaplan, H. J. & Hendrickson, A. E. Development of astrocytes and their relation to blood vessels in fetal monkey retina. *Invest. Ophthalmol. Vis. Sci.* **37**, 2367–2375 (1996).
34. Dorrell, M. I., Aguilar, E. & Friedlander, M. Retinal vascular development is mediated by endothelial filopodia, a preexisting astrocytic template and specific R-cadherin adhesion. *Invest. Ophthalmol. Vis. Sci.* **43**, 3500–3510 (2002).
35. Gerhardt, H. et al. VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163–1177 (2003).
36. Fruttiger, M. et al. PDGF mediates a neuron-astrocyte interaction in the developing retina. *Neuron* **17**, 1117–1131 (1996).
37. Miyawaki, T. et al. Tlx, an orphan nuclear receptor, regulates cell numbers and astrocyte development in the developing retina. *J. Neurosci.* **24**, 8124–8134 (2004).
38. West, H., Richardson, M. W. D. & Fruttiger, M. Stabilization of the retinal vascular network by reciprocal feedback between blood vessels and astrocytes. *Development* **132**, 1855–1862 (2005).
39. Lindblom, P. et al. Endothelial PDGF-B retention is required for proper investment of pericytes in the microvessel wall. *Genes Dev.* **17**, 1835–1840 (2003).
40. Darland, D. C. et al. Pericyte production of cell-associated VEGF is differentiation-dependent and is associated with endothelial survival. *Dev. Biol.* **264**, 275–288 (2003).
41. Shih, S. C. et al. Transforming growth factor β 1 induction of vascular endothelial growth factor receptor 1: mechanism of pericyte-induced vascular survival in vivo. *Proc. Natl Acad. Sci. USA* **100**, 15859–15864 (2003).
42. Antonelli-Orlidge, A., Saunders, K. B., Smith, S. R. & D'Amore, P. A. An activated form of transforming growth factor β is produced by cocultures of endothelial cells and pericytes. *Proc. Natl Acad. Sci. USA* **86**, 4544–4548 (1989).
43. Enge, M. et al. Endothelium-specific platelet-derived growth factor-B ablation mimics diabetic retinopathy. *EMBO J.* **21**, 4307–4316 (2002).
44. Diaz-Araya, C. M., Provis, J. M., Penfold, P. L. & Billson, F. A. Development of microglial topography in human retina. *J. Comp. Neurol.* **363**, 53–68 (1995).
45. Stone, J. et al. Roles of vascular endothelial growth factor and astrocyte degeneration in the genesis of retinopathy of prematurity. *Invest. Ophthalmol. Vis. Sci.* **37**, 290–299 (1996).
46. Vernon, R. B., Angello, J. C., Iruela-Arispe, M. L., Lane, T. F. & Sage, E. H. Reorganization of basement membrane matrices by cellular traction promotes the formation of cellular networks in vitro. *Lab. Invest.* **66**, 536–547 (1992).
47. Rehm, H. L. et al. Vascular defects and sensorineural deafness in a mouse model of Norrie disease. *J. Neurosci.* **22**, 4286–4292 (2002).
48. Xu, Q. et al. Vascular development in the retina and inner ear: control by Norrin and Frizzled-4, a high-affinity ligand-receptor pair. *Cell* **116**, 883–895 (2004).
49. Robitaille, J. et al. Mutant frizzled-4 disrupts retinal angiogenesis in familial exudative vitreoretinopathy. *Nature Gen.* **32**, 326–330 (2002).
50. Niehrs, C. Norrin and frizzled; a new vein for the eye. *Dev. Cell* **6**, 453–454 (2004).
51. Chan-Ling, T. et al. Astrocyte-endothelial cell relationships during human retinal vascular development. *Invest. Ophthalmol. Vis. Sci.* **45**, 2020–2032 (2004).
52. Ohlmann, A. et al. Ectopic norrin induces growth of ocular capillaries and restores normal retinal angiogenesis in Norrie disease mutant mice. *J. Neurosci.* **25**, 1701–1710 (2005).
53. Smith, L. E. et al. Oxygen-induced retinopathy in the mouse. *Invest. Ophthalmol. Vis. Sci.* **35**, 101–111 (1994).
54. Heckenlively, J. R. et al. Mouse model of subretinal neovascularization with choroidal anastomosis. *Retina* **23**, 518–522 (2003).
55. Trommsdorff, M. et al. Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor 2. *Cell* **97**, 689–701 (1999).
56. Dorrell, M. I. et al. Abnormal retinal vascular development associated with a mutation in the gene for the very low density lipoprotein receptor (VLDLR). *Mol. Biol. Cell* **14** (Suppl.), 260 (2003).
57. Grant, M. B. et al. Adult hematopoietic stem cells provide functional hemangioblast activity during retinal neovascularization. *Nature Med.* **8**, 607–612 (2002).
58. Sengupta, N. et al. The role of adult bone marrow-derived stem cells in choroidal neovascularization. *Invest. Ophthalmol. Vis. Sci.* **44**, 4908–4913 (2003).
59. Otani, A. et al. Bone marrow-derived stem cells target retinal astrocytes and can promote or inhibit retinal angiogenesis. *Nature Med.* **8**, 1004–1010 (2002).
60. Otani, A. et al. Rescue of retinal degeneration by intravitreally injected adult bone marrow-derived lineage-negative hematopoietic stem cells. *J. Clin. Invest.* **114**, 765–774 (2004).
61. Dorrell, M. I., Otani, A., Aguilar, E., Moreno, S. K. & Friedlander, M. Adult bone marrow-derived stem cells use R-cadherin to target sites of neovascularization in the developing retina. *Blood* **103**, 3420–3427 (2004).
62. Galimi, F., Summers, R. G., van Praag, H., Verma, I. M. & Gage, F. H. A role for bone marrow-derived cells in the vasculature of noninjured CNS. *Blood* **105**, 2400–2402 (2005).
63. Arfken, C. L., Reno, P. L., Santiago, J. V. & Klein, R. Development of proliferative diabetic retinopathy in African-Americans and whites with type 1 diabetes. *Diabetes Care* **21**, 792–795 (1998).
64. Aiello, L. P. et al. Vascular endothelial growth factor in ocular fluid of patients with diabetic retinopathy and other retinal disorders. *N. Engl. J. Med.* **331**, 1480–1487 (1994).
65. Adamis, A. P. et al. Increased vascular endothelial growth factor levels in the vitreous of eyes with proliferative diabetic retinopathy. *Am. J. Ophthalmol.* **118**, 445–450 (1994).
66. Koyama, R., Nakanishi, T., Ikeda, T. & Shimizu, A. Catalogue of soluble proteins in human vitreous humor by one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis and electrospray ionization mass spectrometry including seven angiogenesis-regulating factors. *J. Chromatogr. B. Anal. Technol. Biomed. Life Sci.* **792**, 5–21 (2003).
67. Hardy, P. et al. New insights into the retinal circulator: inflammatory lipid mediators in ischemic retinopathy. *Prostaglandins Leukot. Essent. Fatty Acids* **72**, 301–325 (2005).
68. Beach, J. M., Schwenzer, K. J., Srinivas, S., Kim, D. & Tiedeman, J. S. Oximetry of retinal vessels by dual-wavelength imaging: calibration and influence of pigmentation. *J. Appl. Physiol.* **86**, 748–758 (1999).
69. Stefansson, E., Machemer, R., de Juan, E., McCuen, B. W. & Peterson, J. Retinal oxygenation and laser treatment in patients with diabetic retinopathy. *Am. J. Ophthalmol.* **113**, 36–38 (1992).
70. Harris, A. et al. Hyperoxia improves contrast sensitivity in early diabetic retinopathy. *Br. J. Ophthalmol.* **80**, 209–213 (1996).
71. Nguyen, Q. D. et al. Supplemental oxygen improves diabetic macular edema: a pilot study. *Invest. Ophthalmol. Vis. Sci.* **45**, 617–624 (2004).
72. Stefansson, E., Hatchell, D. L., Fisher, B. L., Sutherland, F. S. & Machemer, R. Panretinal photocoagulation and retinal oxygenation in normal and diabetic cats. *Am. J. Ophthalmol.* **101**, 657–664 (1986).
73. Stefansson, E., Peterson, J. I. & Wang, Y. H. Intraocular oxygen tension measured with a fiber-optic sensor in normal and diabetic dogs. *Am. J. Physiol.* **256**, H1127–H1133 (1989).
74. Linsenmeier, R. A. et al. Retinal hypoxia in long-term diabetic cats. *Invest. Ophthalmol. Vis. Sci.* **39**, 1647–1655 (1998).
75. Poulaki, V. et al. Insulin-like growth factor-1 plays a pathogenetic role in diabetic retinopathy. *Am. J. Pathol.* **165**, 457–469 (2004).
76. Gardner, T. W., Antonetti, D. A., Barber, A. J., LaNoue, K. F. & Levison, S. W. Diabetic retinopathy: more than meets the eye. *Surv. Ophthalmol.* **47** (Suppl. 2), S253–S262 (2002).
77. Barber, A. J. et al. Neural apoptosis in the retina during experimental and human diabetes. Early onset and effect of insulin. *J. Clin. Invest.* **102**, 783–791 (1998).
78. Amin, R. H. et al. Vascular endothelial growth factor is present in glial cells of the retina and optic nerve of human subjects with nonproliferative diabetic retinopathy. *Invest. Ophthalmol. Vis. Sci.* **38**, 36–47 (1997).
79. Krady, J. K. et al. Minocycline reduces proinflammatory cytokine expression, microglial activation, and caspase-3 activation in a rodent model of diabetic retinopathy. *Diabetes* **54**, 1559–1565 (2005).
80. Holcik, M., Sonenberg, N. & Korneluk, R. G. Internal ribosome initiation of translation and the control of cell death. *Trends Genet.* **16**, 469–473 (2000).

Acknowledgements We thank the following for generously providing material for figures: T. Bennett (Fig. 2b, d), C. Betsholtz (Figs 4c, d, 5c), M. Friedlander (Fig. 6a), M. Fruttiger (Fig. 5d, e), F. Gage (Fig. 6b), F. Galimi (Fig. 6b), A. Otani (Fig. 6a), J. Provis (Fig. 2e) and T. Sandercoe (Fig. 2e). We acknowledge support from the Juvenile Diabetes Research Foundation and the American Diabetes Association (T.G.).

Author Information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to R.F.G. (raygariano@hotmail.com).

Angiogenesis as a therapeutic target

Napoleone Ferrara¹ & Robert S. Kerbel²

Inhibiting angiogenesis is a promising strategy for treatment of cancer and several other disorders, including age-related macular degeneration. Major progress towards a treatment has been achieved over the past few years, and the first antiangiogenic agents have been recently approved for use in several countries. Therapeutic angiogenesis (promoting new vessel growth to treat ischaemic disorders) is an exciting frontier of cardiovascular medicine, but further understanding of the mechanisms of vascular morphogenesis is needed first.

Early pioneers of angiogenic research observed over a century ago that the growth of human tumours is often accompanied by increased vascularity. They suggested that a key aspect of the cancer process is a disease of the vasculature in the whole area affected (reviewed in ref. 1). The existence of tumour-derived factors responsible for promoting new vessel growth was postulated over 65 years ago², and a few years later it was proposed that tumour growth is crucially dependent on the development of a neovascular supply³. In 1971, it was hypothesized that inhibition of angiogenesis (antiangiogenesis) would be an effective strategy to treat human cancer, and an active search for angiogenesis inducers and inhibitors began⁴. Extensive research has led to the identification and isolation of several regulators of angiogenesis, some of which represent therapeutic targets.

Despite some initial setbacks and negative clinical trial results, major progress has been made over the past few years in targeting angiogenesis for human therapy. In February 2004, the US Food and Drug Administration (FDA) approved bevacizumab, a humanized anti-VEGF (vascular endothelial growth factor)-A monoclonal antibody, for the treatment of metastatic colorectal cancer in combination with 5-fluorouracil (FU)-based chemotherapy regimens. This fol-

lowed from a phase III study showing a survival benefit⁵. In December 2004, the FDA approved pegaptinib, an aptamer that blocks the 165 amino-acid isoform of VEGF-A, for the treatment of the wet (neovascular) form of age-related macular degeneration (AMD)⁶.

These achievements have validated the notion that angiogenesis is an important target for cancer and other diseases. These advances notwithstanding, much progress is needed on a variety of important issues; for example, how do we achieve the most effective combinations of antiangiogenic agents with chemotherapy or other biological agents and how do we select patients that are most likely to respond to the treatment? Another issue is that resistance to antiangiogenic therapy is emerging⁷ and thus a better understanding of pathways that may mediate tumour angiogenesis in various circumstances is necessary. Furthermore, the hope that 'therapeutic angiogenesis' will provide a treatment for ischaemic disorders still remains unfulfilled, in spite of considerable preclinical and clinical efforts.

The main purpose of this review is to summarize recent progress and emphasize the issues that need to be resolved before the field of angiogenic therapy can make further significant advances.

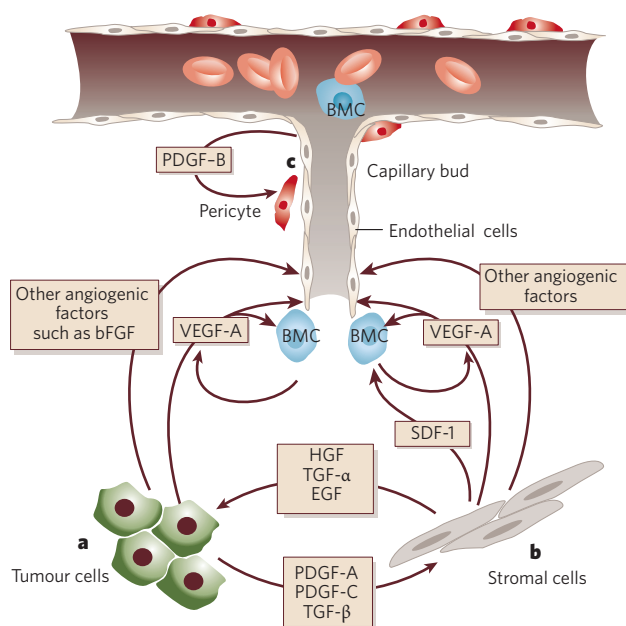


Figure 1 | A few of the molecular and cellular players in the tumour/microvascular microenvironment. a, Tumour cells produce VEGF-A and other angiogenic factors such as bFGF, angiopoietins, interleukin-8, PlGF and VEGF-C. These stimulate resident endothelial cells to proliferate and migrate. **b,** An additional source of angiogenic factors is the stroma. This is a heterogeneous compartment, comprising fibroblastic, inflammatory and immune cells. Recent studies indicate that tumour-associated fibroblasts produce chemokines such as SDF-1, which may recruit bone-marrow-derived angiogenic cells (BMC). The various hypotheses on the nature and role of such cells in angiogenesis and tumour progression are discussed in the text. VEGF-A or PlGF may also recruit BMC. Tumour cells may also release stromal cell-recruitment factors, such as PDGF-A, PDGF-C or transforming growth factor (TGF)-β. A well-established function of tumour-associated fibroblasts is the production of growth/survival factor for tumour cells such as EGFR ligands, hepatocyte growth factor and heregulin. **c,** Endothelial cells produce PDGF-B, which promotes recruitment of pericytes in the microvasculature after activation of PDGFR-β. HGF, hepatocyte growth factor.

¹Genentech, 1 DNA Way, South San Francisco, California 94080, USA; ²Sunnybrook and Women's College Health Sciences Centre and the University of Toronto, Ontario M5G 2M9, Canada.

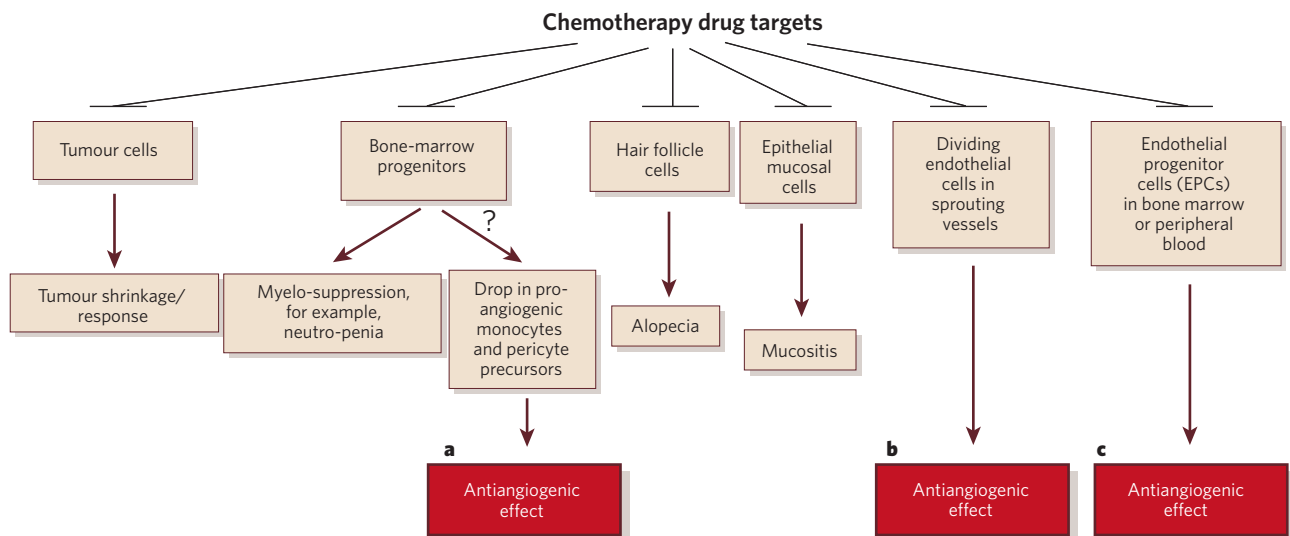


Figure 2 | Chemotherapy targets. In addition to tumour cells, the intended target for chemotherapy in cancer patients, conventional chemotherapy drugs can inhibit the proliferation of, or kill, a number of normal host cell types, including several that, in principle, can contribute to an antiangiogenic effect. Targeting of various normal cell populations is generally associated with harmful or undesirable side effects such as myelosuppression, alopecia or mucositis. A desirable effect could be antiangiogenesis as a result of targeting. **a**, Bone-marrow-derived proangiogenic cells that adhere to the walls of new blood vessels and further stimulate their growth by paracrine mechanisms. Whether these latter cell types, which probably include monocytes and pericyte precursors, are affected directly by chemotherapy or are reduced in numbers by elimination of more primitive bone marrow progenitors which give rise to such cells is not yet clearly established. **b**, Cycling endothelial cells present in sprouting blood vessel capillaries; and **c**, authentic bone-marrow-derived circulating endothelial progenitor cells (EPC) that can incorporate into the lumen of growing vessels and differentiate into endothelial cells. Inhibiting the levels or function of VEGF can augment these various antiangiogenic mechanisms of chemotherapy. For example, VEGF is a potent mobilizer of EPC, a pro-survival (anti-apoptotic) factor for differentiated, activated endothelial cells, and also may be one of the more important paracrine growth factors secreted by proangiogenic vessel adherent bone-marrow-derived monocytes.

The major signalling pathways in tumour angiogenesis VEGF/VEGF receptors

Angiogenesis is a fundamental developmental and adult physiological process, requiring the coordinated action of a variety of growth factors and cell-adhesion molecules in endothelial and mural cells (reviewed in this issue by Coultas, Chawengsaksophak and Rossant, p. 937). So far, VEGF-A and its receptors are the best-characterized signalling pathway in developmental angiogenesis^{1,8,9}. Loss of a single VEGF-A allele results in embryonic lethality^{1,8,9}. This pathway also has an essential role in reproductive and bone angiogenesis⁸. Much research has also established the role of VEGF-A in tumour angiogenesis^{8,10}. VEGF-A binds to two receptor tyrosine kinases (RTK), VEGFR-1 (Flt-1) and VEGFR-2 (KDR, Flk-1) (reviewed in ref. 10). Of the two, it is now generally agreed that VEGFR-2 is the major mediator of the mitogenic, angiogenic and permeability-enhancing effects of VEGF-A. The significance of VEGFR-1 in the regulation of angiogenesis is more complex. Under some circumstances, VEGFR-1 may function as a 'decoy' receptor that sequesters VEGF and prevents its interaction with VEGFR-2 (ref. 10). However, there is growing evidence that VEGFR-1 has significant roles in haematopoiesis and in the recruitment of monocytes and other bone-marrow-derived cells that may home in on the tumour vasculature and promote angiogenesis^{11–13}. In addition, VEGFR-1 is involved in the induction of matrix metalloproteinases (MMPs)¹⁴ and in the paracrine release of growth factors from endothelial cells¹⁵. Thus the VEGFR-1-selective ligands VEGF-B and placental-like growth factor (PLGF) may also have a role in these processes. Furthermore, in some cases VEGFR-1 is expressed by tumour cells and may mediate a chemotactic signal, thus potentially extending the role of this receptor in cancer growth¹⁶.

VEGF-A gene expression is upregulated by hypoxia¹⁷. The transcription factor hypoxia inducible factor (HIF), which operates in concert with the product of the von Hippel-Lindau (VHL) tumour suppressor gene, has a major role in such regulation. Under normoxic conditions, the VHL protein targets HIF for ubiquitination and degradation¹⁷.

In situ hybridization studies demonstrate that VEGF-A messenger

RNA is expressed in many human tumours¹⁸. Renal cell carcinomas have a particularly high level of VEGF-A expression, consistent with the notion that inactivating VHL mutations occur in about 50% of such tumours¹⁹, thus providing a further explanation for the responsiveness of this tumour type to a VEGF-A blockade²⁰. However, VEGF-A upregulation in tumours is not only linked to hypoxia or VHL mutations. Indeed, a very broad and diverse spectrum of oncogenes is associated with VEGF-A upregulation, including mutant ras, erbB-2/Her2, activated EGFR and bcr-abl^{7,21}. Besides VHL, inactivation/mutation of various other suppressor genes can also result in VEGF upregulation. These genes include those associated with familial syndromes characterized by well-vascularized hamartomas²².

In 1993, it was reported that a murine anti-human VEGF-A monoclonal antibody inhibited the growth of several tumour cell lines in nude mice, whereas the antibody had no effect on tumour-cell proliferation *in vitro*²³. Subsequent studies have shown that many additional tumour cell lines, regardless of the tumour's origin, are inhibited *in vivo* by the same anti-VEGF monoclonal antibody (reviewed in ref. 24). Tumour-growth inhibition has also been demonstrated using independent anti-VEGF approaches including a dominant-negative VEGFR-2 mutant²⁵, anti-VEGFR-2 antibodies²⁶, small molecule inhibitors of VEGF RTKs²⁷ and soluble VEGF receptors^{28,29}. VEGF-A gene inactivation also suppresses angiogenesis in a transgenic model of multi-stage tumorigenesis³⁰.

Platelet-derived growth factor (PDGF) and angiopoietins

Other signalling molecules that have an established role in the development and differentiation of the vessel wall such as PDGF-B/PDGFR- β ³¹ and the angiopoietins (Ang), the ligands of the Tie2 receptor⁹, may also be therapeutic targets. PDGF-B is required for recruitment of pericytes and maturation of the microvasculature³¹. Inhibition of PDGFR- β signalling has been reported to result in a tumour microvascular tree that is particularly dependent on VEGF-mediated survival signals. Withdrawal of VEGF-A leads to endothelial apoptosis and vascular regression³². In this context, newly formed

vessels, whether they are tumour-associated or not, are particularly vulnerable to VEGF-A blockade, whereas mature vessels, covered by extracellular matrix and pericytes, may be resistant to VEGF inhibitors and other antiangiogenic agents. Furthermore, recent studies have emphasized the significance of tumour-derived PDGF-A (and potentially PDGF-C) and PDGFR- α signalling in the recruitment of an angiogenic stroma that produces VEGF-A and other angiogenic factors³³ (Fig. 1). Therefore, combining PDGF and VEGF inhibitors is an attractive anti-vascular and anti-tumour strategy.

Ang-1 is required for further remodelling and maturation of the initially immature vasculature. Unlike mouse embryos lacking VEGF-A or VEGFR-2, embryos lacking Ang-1 or its receptor Tie2 develop a rather normal primary vasculature, but this vasculature fails to undergo effective remodelling (reviewed in ref. 9). The generally accepted view is that Ang-1 is the major agonist for Tie2, whereas Ang-2 may act as an antagonist or a partial agonist³⁴. However, more recent evidence indicates that, unexpectedly, Ang-2 has a positive role, at least in tumour angiogenesis³⁵. Administration of Ang-2 inhibitors to tumour-bearing mice has been reported to result in delayed tumour growth, accompanied by reduced endothelial cell proliferation, consistent with an antiangiogenic mechanism. Therefore, inhibitors of Ang-2 may be candidates for clinical development³⁵.

Axon-guidance molecules

Recently, the role of axon-guidance receptors and ligands in developmental angiogenesis has received much attention. There are four main families: the neuropilins (NRP)/semaphorins, the ephrins, Robo/Slit and netrin/Unc5. For recent reviews, see refs 36, 37. Although the significance of these pathways in tumour angiogenesis is far from clear, there is emerging evidence that they have a role in some cancer models and therefore may be potential therapeutic targets.

NRP1 and NRP2, previously shown to bind the collapsin/semaphorin family and implicated in axon guidance, are also receptors for the heparin-binding isoforms of VEGF-A and seem to potentiate the activation of VEGFR-2 by VEGF165 (ref. 38). Therefore, NRPs may participate in tumour angiogenesis as positive modulators of VEGF signalling in endothelial cells. Furthermore, NRP1 and NRP2 are expressed on the cell surface of several tumour cell lines that bind VEGF165 and display a chemotactic response to this ligand, suggesting a pro-tumour activity of NRPs, with or without the involvement of VEGF RTK signalling³⁷.

The ephrins and their tyrosine kinase Eph receptors are a large fam-

ily, initially implicated in neuronal guidance during development and subsequently found to have activities in other cell types, including vascular cells (for a review see ref. 39). The earliest evidence for a role of this family in angiogenesis was the report by Pandey *et al.* that ephrin A1 mediates TNF- α -induced angiogenesis *in vivo*⁴⁰. Ephrin B2 and its receptor EphB4 are important for distinguishing between developing arterial and venous vessels (see p. 937). Recent studies suggest a role for Eph/ephrin interactions in malignant tumour progression and angiogenesis. Soluble EphB4-expressing human melanoma A375 cells grown subcutaneously in nude mice showed reduced tumour growth compared with control tumours⁴¹. Interfering with EphA signalling has been also reported to result in some inhibition of angiogenesis in tumour models⁴².

Slits are secreted proteins that function as chemorepellents in axon guidance and neuronal migration through the Roundabout (Robo) receptor (reviewed in refs 36, 37). Wang *et al.* reported the expression of Slit2 in several tumour cell types and that Robo1 expression was localized to vascular endothelial cells⁴³. Recombinant Slit2 protein attracted endothelial cells and promoted tube formation. Neutralization of Robo1 reduced microvessel density and growth of A375 cells transplanted in nude mice⁴³.

Negative regulators of angiogenesis

Angiogenesis is a tightly regulated process and seems to be under the control of both positive and negative regulatory factors. Although several potential negative regulators of angiogenesis have been identified, relatively little is known about their role in the physiological regulation of angiogenesis. Thrombospondin, a large multifunctional glycoprotein secreted by most epithelial cells in the extracellular matrix, inhibits angiogenesis associated with tumour growth and metastasis⁴⁴. Several fragments of larger proteins have been described as endogenous inhibitors of angiogenesis including endostatin⁴⁵, tumstatin⁴⁶ and vasostatin⁴⁷. The most recently described endogenous inhibitor of angiogenesis is vasohibin, which seems to be derived from the endothelium and to operate as a feedback regulator⁴⁸. The precise mechanism of action of these proteins remains to be more clearly defined, although several hypotheses have been proposed, including that they bind to specific integrins in the case of endostatin and tumstatin⁴⁹.

Role of bone-marrow-derived cells in angiogenesis

An intensively debated issue in the field is the contribution (as well as the precise nature) of bone-marrow-derived endothelial progenitor

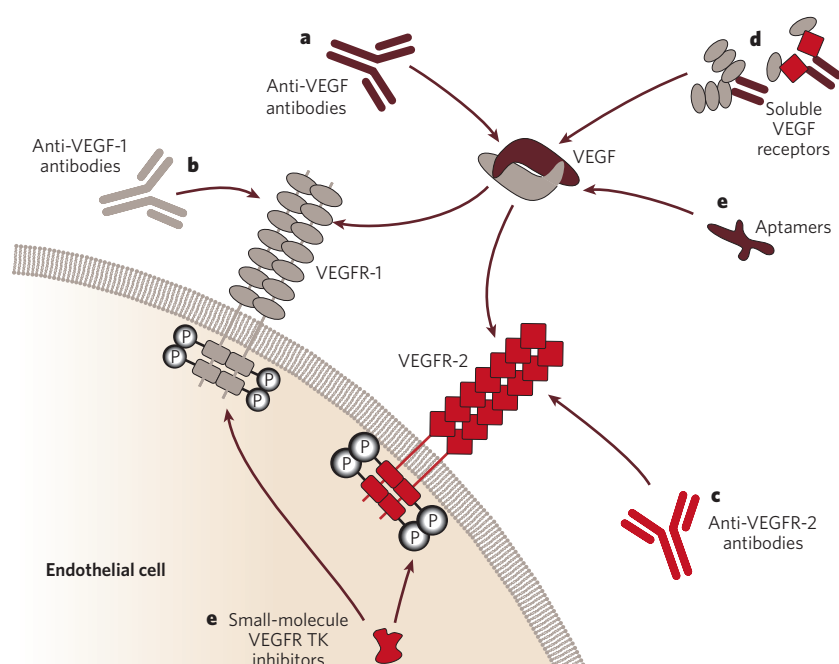


Figure 3 | Various strategies to inhibit VEGF signalling. These include monoclonal antibodies targeting VEGF-A (a) or the VEGF receptors (b, c). d, Chimaeric soluble receptors such as the 'VEGF-trap' (domain 2 of VEGFR-1 and domain 3 of VEGFR-2 fused to a Fc fragment of an antibody) are also undergoing clinical development. e, Additional extracellular inhibitors are aptamers that bind the heparin-binding domain of VEGF165 (pegaptanib). A variety of small-molecule VEGF RTK inhibitors that inhibit ligand-dependent receptor autophosphorylation of VEGFR-1 and VEGFR-2 are being tested. Additional strategies to inhibit VEGF signalling include antisense and siRNA targeting VEGF-A or its receptors.

cells (EPC) to angiogenesis. However, there is little doubt that bone-marrow-derived cells participate in angiogenic processes, at least as a source of angiogenic factors. In 1997, Asahara *et al.* reported the isolation of putative EPC from human peripheral blood, on the basis of cell-surface expression of CD34 and other endothelial markers⁵⁰. These cells were reported to differentiate *in vitro* into endothelial cells and seemed to be incorporated at sites of active angiogenesis in various animal models of ischaemia. These findings suggested that incorporation into the lumen of bone-marrow-derived endothelial precursor cells contributes to the growing vessels, complementing resident endothelial cells in sprouting new vessels. Also, ischaemia and various cytokines, including VEGF and granulocyte-macrophage colony-stimulating factor (GM-CSF), were reported to mobilize EPC into sites of neovascularization⁵¹. However, the precise contribution of these cells in various pathophysiological circumstances was not clearly defined.

Subsequent studies have suggested that the contribution of such cells to angiogenesis is dependent on the experimental system employed. In the angiogenic-defective, tumour-resistant *Id*-mutant mice, EPC accounted for a large proportion of endothelial cells associated with xenografted tumours⁵². Rafii and collaborators proposed that mobilization of EPC from bone marrow requires angiogenic-factor-mediated activation of MMP-9, which leads to the release of the soluble KIT ligand. This ligand would in turn promote proliferation and motility of EPC within the bone-marrow microenvironment, thus creating permissive conditions for their mobilization into the peripheral circulation⁵³. However, in spontaneous tumours occurring in *Id*-deficient mice in the tumour-prone PTEN^{+/−} genetic background, the contribution of EPC was less significant⁵⁴. Also, De Palma *et al.* suggested that the percentage of EPC that are truly incorporated into a growing vessel wall is very low and that the majority of bone-marrow-derived cells homing in on the tumour vasculature are adherent perivascular mononuclear cells, which may contain angiogenic factors⁵⁵. Peters *et al.* recently analysed the tumour endothelial cells in six individuals who developed cancers after bone-marrow transplantation with donor cells derived from individuals of the opposite sex and found that an average of 4.9% of cells of the total endothelial cell population were derived from the bone marrow⁵⁶.

In summary, bone-marrow-derived cells seem to contribute to tumour angiogenesis, of which a small and variable proportion are probably true EPCs. Bone-marrow-derived circulating pro-angiogenic cells, regardless of their precise nature, may be a common target for antiangiogenic therapies and may be exploitable as surrogate biomarkers for the angiogenic process as well as antiangiogenic therapies⁵⁷.

Combination therapies

It is increasingly likely that cancer therapy, with a few exceptions, will need to be combinatorial. It seems logical to target multiple pathways simultaneously. Much preclinical evidence indicates that combining antiangiogenic agents with conventional cytotoxic agents or radiation therapy results in additive or even synergistic anti-tumour effects⁵⁸. So far, it is unclear whether such positive interaction takes place preferentially with specific types of antiangiogenic or cytotoxic agents. An issue that is being debated is the mechanism of such potentiation, as it would seem counterintuitive that 'tumour-starving' antiangiogenic drugs that suppress blood flow in tumours actually increase the efficacy of chemotherapy. Browder *et al.*⁵⁹ and Klement *et al.*⁶⁰ proposed that chemotherapy, especially when delivered at close regular intervals using relatively low doses with no prolonged drug-free break periods ('metronomic therapy'), preferentially damages endothelial cells in tumour blood vessels. These cells are presumably dividing, and the simultaneous blockade of VEGF-A is thought to blunt a key survival signal for endothelial cells, thus selectively amplifying the endothelial cell targeting effects of chemotherapy, leading to improved subsequent killing of cancer cells.

A similar process, in principle, may take place when combining

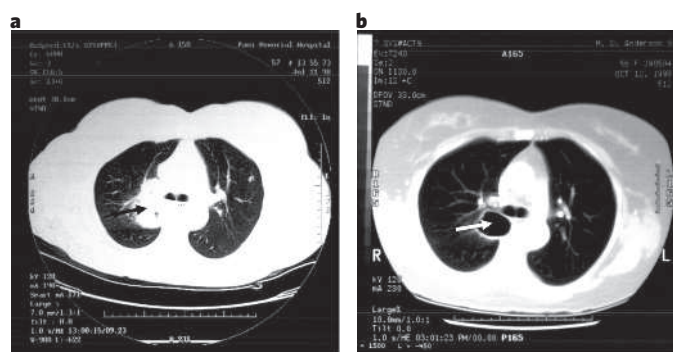


Figure 4 | Computed tomography chest scans. These scans were taken of a NSCLC patient before (a) and after (b) three cycles (nine weeks) of treatment with bevacizumab plus carboplatin and taxol (reproduced from ref. 71 with permission from American Society of Clinical Oncology). Note that the tumour mass in a (arrow) underwent extensive necrosis and cavitation in b (arrow). This pattern was seen more frequently in patients treated with bevacizumab plus chemotherapy relative to chemotherapy alone. Cavitation may be associated with serious bleeding, especially when it occurs in proximity to large vessels⁷¹.

more conventional maximum-tolerated dose chemotherapy regimens with a drug such as bevacizumab⁶¹. In addition, bone-marrow-derived pro-angiogenic circulating cells, probably including authentic EPC, seem to be very sensitive to both conventional cytotoxic and low-dose metronomic chemotherapy⁶². However, levels of such cells can rapidly rebound, returning to normal or even increased levels during the drug-free break periods after maximum-tolerated dose cytotoxic chemotherapy⁶². Because VEGF-A acts as a mobilizing and probably a survival agent for such cells, co-administration of a VEGF-targeting agent, especially one with a long half-life in the circulation (for example, anti-VEGF antibodies), would be expected to amplify and sustain the suppressive effects of standard (as well as metronomic) chemotherapy on bone-marrow-derived circulating pro-angiogenic cells⁶³. Furthermore, it has been proposed that progressively accelerated proliferation and repopulation of cancer cells during intervals of radiotherapy or chemotherapy is an important cause of treatment failure⁶⁴. It is tempting to speculate that antiangiogenic treatment during these intervals inhibits such repopulation process. Figure 2 illustrates the various cellular targets of chemotherapy.

An alternative hypothesis has been proposed by Jain⁶⁵. Submaximal doses of an antiangiogenic agent such as an anti-VEGFR-2 antibody would 'normalize' the vasculature that is characteristic of many vessels in tumours. This would result in pruning of excessive endothelial and perivascular cells, in a decrease in the normally high interstitial pressures detected in solid tumours and in temporarily improved oxygenation and delivery of chemotherapy to tumour cells⁶⁵. However, according to recent studies, the tumour vasculature can be 'normalized' transiently, and eliciting synergistic effects through this mechanism requires administration of chemotherapy or radiation therapy over a defined time window after the angiogenesis inhibitor⁶⁶. Considering also that in most clinical protocols no such intentional sequential administration is performed, it remains to be established whether such a mechanism accounts for the long-term beneficial effects of combination treatments observed in some trials. By contrast, acute administration of angiogenic inhibitors induces vascular changes consistent with 'normalization' in humans. In this regard, Willett *et al.* reported that a single infusion of bevacizumab to patients with rectal carcinoma rapidly decreased tumour perfusion, vascular volume, microvascular density and interstitial fluid pressure as well as the number of viable, circulating endothelial cells in six colorectal cancer patients⁶⁷.

Combinatorial therapies with antiangiogenic agents are not limited to those including cytotoxic chemotherapy. Several preclinical and clinical trials are exploring the combination of various angiogenesis

inhibitors with other targeted therapies, such as EGFR or Her2 inhibitors (cetuximab, erlotinib and trastuzumab), PDGFR/ bcr-abl inhibitors (imatinib), proteasome inhibitors (bortezomib) and other antiangiogenic agents such as inhibitors of integrins (for example $\alpha v\beta 3$ and $\alpha 5\beta 1$).

Clinical trials for antiangiogenesis

Many angiogenesis inhibitors are currently in clinical trials. It is noteworthy that, in parallel to angiogenesis inhibitors, another class of vascular-targeting or vascular-modulating drugs is being tested, namely 'vascular-disrupting agents'. These drugs primarily target existing, recently formed vasculature and cause acute vascular occlusion and disruption of tumour blood flow⁶⁸. For an overview of these trials, see <http://www.cancer.gov/clinicaltrials/developments/anti-angio-table>. The inhibitors tested include a variety of agents with diverse mechanisms of action (several of which are not known). At present, inhibitors of the VEGF pathway are the most clinically advanced, and bevacizumab, a humanized variant of a murine anti-VEGF-A monoclonal antibody that was used in early proof-of-concept studies²³, is the only FDA-approved antiangiogenic treatment for cancer therapy⁶⁹. Figure 3 illustrates several methods for inhibiting the VEGF pathway.

Several important clinical studies testing angiogenesis inhibitors have been presented at recent oncology meetings, such as the American Society of Clinical Oncology meeting. Typically, clinical studies are presented and discussed at such meetings in advance of peer-reviewed publication. Therefore, in the interest of an up-to-date overview of the field, a discussion of some of these studies will be included here, with the caveat that the data are preliminary and require further analysis.

The clinical trial that resulted in FDA approval of bevacizumab was a large, randomized, double-blind, phase III study in which bevacizumab was administered in combination with bolus IFL (irinotecan, 5FU and leucovorin) chemotherapy as first-line therapy for metastatic colorectal cancer⁵. Median survival was increased from 15.6 months in the bolus-IFL plus placebo arm of the trial to 20.3 months in the bolus IFL plus bevacizumab arm. Similar increases were seen in progression-free survival, response rate and duration of response. The clinical benefit of bevacizumab was seen in all subject subgroups, including those defined by performance status, location of primary tumour, number of organs involved and duration of metastatic disease⁵. Although bevacizumab was generally well tolerated, some serious and unusual toxicities have been noted, albeit at low frequencies. Bevacizumab was associated with gastrointestinal perforations and wound healing complications in about 2% of patients. In addition, the incidence of arterial thromboembolic complications were increased about twofold relative to chemotherapy alone, with patients 65 years or older with a history of arterial thromboembolic events being at higher risk. Although the precise mechanism of this effect is unknown, it is conceivable that vascular damage induced by cytotoxic agents can be exacerbated by the blockade of VEGF-A.

Preliminary data of a phase III study indicate that bevacizumab confers a survival advantage on patients with previously treated, relapsed, metastatic colorectal cancer in combination with FOLFOX4 chemotherapy (5-fluorouracil, leucovorin and oxaliplatin), relative to chemotherapy alone (B. Giantonio, P. J. Catalano, N. J. Meropol, E. P. Mitchell, M. A. Schwartz *et al.*, unpublished data).

The role of bevacizumab in other tumour types and settings is currently under investigation, and phase III clinical trials of this drug in non-small-cell-lung cancer (NSCLC), renal cell cancer and metastatic breast cancer are ongoing. An early phase III trial of advanced, heavily pretreated, metastatic breast cancer showed that adding bevacizumab to capecitabine chemotherapy did not improve progression-free survival, despite a doubling of the response rate (that is, tumour shrinkage of 50% or more) in the bevacizumab-treated arm of the trial⁷⁰. Thus, the responses seemed to be very short in duration. However, an interim analysis of a phase III study of women with previously untreated metastatic breast cancer treated with bevacizumab in com-

bination with weekly paclitaxel chemotherapy showed that the study met its primary efficacy endpoint of improving progression-free survival, compared with paclitaxel alone (K. Miller, unpublished data).

Furthermore, administration of bevacizumab in combination with paclitaxel and carboplatin to patients with NSCLC resulted in increased response rate and time to progression relative to chemotherapy alone in a randomized phase II trial⁷¹. The most significant adverse event was serious haemoptysis. This was primarily associated with centrally located tumours with squamous histology, cavitation and central necrosis and proximity of disease to large vessels⁷¹. Figure 4 illustrates the extensive tumour necrosis and cavitation that may result from the combination treatment⁷¹. More recently, preliminary results from a large, randomized phase III clinical trial for patients with previously untreated advanced non-squamous NSCLC show that patients who received bevacizumab in combination with paclitaxel and carboplatin lived longer than patients who received chemotherapy alone (A. B. Sandler, R. Gray, J. Brahmer, A. Dowlati, J. H. Schiller *et al.*, unpublished data). Serious bleeding was infrequent but occurred more commonly in the bevacizumab arm of the trial.

Besides bevacizumab, several other VEGF inhibitors are being clinically pursued. A variety of small-molecule RTK inhibitors targeting the VEGF receptors have been developed. The most advanced are SU11248 and Bay 43-9006. SU11248 inhibits VEGFRs, PDGFR, c-kit and Flt-3 (ref. 72) and has been reported to have considerable efficacy in imatinib-resistant gastrointestinal stromal tumour (R. G. Maki, J. A. Fletcher, M. C. Heinrich, J. A. Morgan, S. George *et al.*, unpublished data). Bay 43-9006 was initially identified as a raf kinase inhibitor and subsequently shown to inhibit several RTKs including VEGFRs. An interim analysis of phase III data indicates that Bay 43-9006 monotherapy results in a significant increase in progression-free survival in patients with advanced renal cell carcinoma (B. Escudier, C. Szczylik, T. Eisen, W. M. Stadler, B. Schwartz *et al.*, unpublished data). Follow-up of such phase III data is ongoing to determine whether an overall survival benefit occurs. An earlier randomized phase II study had shown that bevacizumab as a single agent also results in an increase in time to progression in renal cell carcinoma patients, providing further evidence that this tumour type may be particularly responsive to anti-VEGF treatment²⁰.

AG-013736, which has a similar spectrum of kinase inhibition to SU11248, has also shown promise in metastatic renal cell carcinoma in a phase II monotherapy study (B. Rini, O. Rixe, R. Bukowski, M. D. Michaelson, G. Wilding *et al.*, unpublished data). Twenty-four patients (46% of those on the trial) experienced partial responses to the treatment. Stable disease was observed in an additional 21 patients (40%). These interesting efficacy data will need to be followed by an appropriately designed and powered phase III trial.

An additional VEGF RTK inhibitor in late-stage clinical trials is PTK787 (ref. 27). This molecule is in phase III trials in colorectal cancer patients, in combination with FOLFOX4 chemotherapy. Recently, interim findings of this trial have been presented (J. R. Hecht, T. Traub-Dietatz, E. Jaeger, J. Hainsworth, R. Wolff *et al.*, unpublished data). According to investigator-based assessment, there was a statistically significant increase in progression-free survival in PTK787-treated patients. However, a central review failed to document any significant difference. Subgroup analysis suggested that patients with high lactic dehydrogenase have the best response to PTK787 in terms of progression-free survival.

A chimaeric, soluble VEGF receptor ('VEGF-trap')²⁹ is also undergoing clinical development as an anti-cancer agent and preliminary results of a phase I study have been recently presented (J. Dupont, M. L. Rothenberg, D. R. Spriggs, J. M. Cedarbaum, E. S. Furfine *et al.*, unpublished data).

Recombinant human endostatin has been tested in phase I studies over the past few years to determine its safety and pharmacokinetic characteristics in patients with solid tumours, and these studies have documented a lack of dose-limiting toxicity⁷³.

Neovascularization and vascular leakage are a major cause of visual

loss in the wet form of AMD⁷⁴. Currently, several anti-VEGF strategies are being explored in AMD clinical trials. The most clinically advanced inhibitors are pegaptanib, an aptamer that interacts with the heparin-binding domain of VEGF165 (ref. 75) and ranibizumab, a humanized high-affinity anti-VEGF Fab that neutralizes all VEGF-A isoforms and proteolytic fragments⁷⁶. Pegaptanib has been approved by the FDA for the treatment of AMD, after phase III studies showing that intraocular administrations of the drug reduced visual loss relative to the placebo⁶. Very recently, data of a controlled phase III study with ranibizumab indicated that the treatment is efficacious and maintains or improves vision in patients with wet AMD (P. J. Rosenfeld, D. M. Browne, J. S. Heier, D. S. Boyer & P. K. Keiser *et al.*, unpublished data).

Resistance to antiangiogenic therapies

Many patients treated with VEGF inhibitors, especially in combination with chemotherapy, may survive longer, but they eventually succumb to their disease. There is emerging evidence that VEGF-A may be replaced by other angiogenic pathways as the disease progresses. These potentially include VEGF-unrelated pathways and others mediated by other members of the VEGF gene family such as the lymphangiogenic factors VEGF-C and VEGF-D, which may bind to and activate VEGFR-2 after proteolytic cleavage (see the review in this issue by Alitalo, Tammela and Petrova, p. 946). Other possible mechanisms for acquired resistance to antiangiogenic drugs^{77,78} include selection and overgrowth of tumour-cell variants that are 'hypoxic resistant' and thus less dependent on angiogenesis⁷⁹ and tumour-vessel remodelling resulting in a shift to mature, stabilized vessels that are less responsive to antiangiogenic drugs⁸⁰. The recruitment of bone-marrow-derived angiogenic cells may also provide a potential mechanism of escape to some antiangiogenic strategies.

Orimo *et al.* have recently reported that one of the mechanisms by which cancer-associated fibroblasts contribute to tumour angiogenesis is the release of stromal cell-derived factor (SDF)-1, which leads to the recruitment of bone-marrow cells in the tumour vasculature (Fig. 1)⁸¹. A current focus of research is the role of myeloid cells of the monocyte/macrophage lineage in mediating multiple pathways leading to tumour progression and angiogenesis (reviewed in ref. 82). The hypoxic tumour microenvironment may lead to upregulation of chemokines and other factors (for example, interleukin-8, angiopoietins, basic fibroblast growth factor (bFGF) and hepatocyte growth factor (HGF) from tumour-infiltrating macrophages, which may facilitate angiogenesis, invasion and immune evasion of tumours⁸². Further potential mechanisms of escape are integrin-signalling-mediated events, as there is growing evidence to support positive interactions between integrins and a variety of RTKs involved in tumour invasion, metastasis and angiogenesis⁸³.

Furthermore, recent studies suggest that in some cases endothelial cells associated with tumours are not a genetically stable, non-transformed, compartment (as long assumed) and instead may provide a further mechanism of resistance to antiangiogenic therapies. Hida *et al.* described cytogenetical abnormalities in murine endothelial cells isolated from human tumour xenografts (aneuploidy, multiple abnormal centrosomes), although the mechanism for the acquisition of such alterations remains unknown⁸⁴. Also, Streubel *et al.* reported that a significant percentage of the endothelial cells in human B-cell lymphomas harbour lymphoma-specific chromosomal abnormalities, suggesting that endothelial cells in B-cell lymphomas may be in part tumour related⁸⁵.

Pro-angiogenic therapies

The possibility that pharmacological treatments for disorders characterized by inadequate tissue perfusion may become available is attractive, as often there are no effective alternatives to surgical reconstruction procedures. Many angiogenic factors are active in a variety of animal models of coronary or limb ischaemia (reviewed in refs 86, 87). On the basis of these preclinical data, several angiogenic

factors (VEGF-A, VEGF-C, FGF-1, FGF-2 and FGF-4) have been tested in patients with myocardial or limb ischaemia⁸⁷. Although promising results were initially reported in small open-label trials, none of the placebo-controlled trials so far conducted with recombinant proteins or gene therapy has yielded convincingly positive results^{88,89}. The placebo effect in these trials proved to be considerably greater than initially suspected. It is conceivable that young and otherwise healthy animals are able to mount an effective endogenous angiogenic response to vascular injury that can be further augmented. However, endothelial cells in patients with extensive atherosclerotic disease may be in part refractory to angiogenic factors^{86,87}. Interestingly, more recent studies suggest that a therapy using a 'master switch' such as HIF might provide advantages relative to single angiogenic factors⁹⁰. A more persistent exposure to angiogenic factors than that achieved in early trials may also lead to better results.

Transplantation of angiogenic cells from bone marrow or peripheral blood to treat ischaemic disorders, as an alternative to angiogenic factors, has also generated significant interest⁹¹. Different strategies have been employed, including the use of bone-marrow cells and either whole peripheral mononuclear cells or purified CD34⁺ cells, after bone-marrow stimulation with cytokines such as GM-CSF⁹¹. The first randomized study testing this hypothesis consisted of an intramuscular injection of autologous bone-marrow-derived mononuclear cells in patients with limb ischaemia⁹². This therapy promoted vascular proliferation and resulted in a significant increase in transcutaneous oxygen pressure, reduced rest pain and increased pain-free walking time in 22 patients with leg ischaemia, relative to injection of peripheral mononuclear cells⁹². Additional small trials in coronary or limb ischaemia patients have reported some improvement in angina and walking time after transplantation of blood cells⁸⁹. Large placebo-controlled studies are required to verify these findings and assess any long-term benefit.

These results highlight a concern regarding possible delayed side effects of long-term antiangiogenic therapy. If bone-marrow-derived cells contribute significantly to vessel integrity and repair, can we expect increases in certain adverse cardiovascular events in cancer patients — especially in older patients with subclinical cardiovascular disease — who are successfully treated with antiangiogenic therapies over prolonged periods?

Conclusions and perspectives

There is now proof that an antiangiogenic approach, when combined with chemotherapy, results in increased survival in patients with advanced malignancies. It could be argued that the clinical benefit and impact would be greater if the therapy were initiated at earlier stages of malignancy, possibly in the adjuvant setting, in agreement with many preclinical data. However, lengthy clinical trials will be required to test this hypothesis. Very recent studies suggest that there is reason to be optimistic and that such studies are warranted. A targeted therapy such as trastuzumab given in combination with chemotherapy to patients with early stage Her2-positive breast cancer results in a 52% reduction in the risk of disease recurrence, compared with those patients who received chemotherapy alone, after four years in the study⁹³. This is a substantial reduction in the risk of cancer recurrence.

It is important to obtain reliable markers to monitor the activity of antiangiogenic drugs. So far, the absence of such biomarkers may have impaired clinical development of various antiangiogenic drugs. Circulating endothelial cells and their progenitor subset and MRI dynamic measurement of vascular permeability/flow in response to angiogenesis inhibitors are potential candidates, although their long-term predictive value remains to be established^{157,67,94}.

Inhibition of lymphangiogenesis has been proposed as a strategy to reduce lymphatic metastasis. Work over the past few years has elucidated the role of the VEGF-C/VEGFR-3 pathway in the regulation of normal and tumour-associated lymphangiogenesis (see p. 946). Therefore, combining traditional antiangiogenic agents with inhibitors of lymphangiogenesis may provide a powerful anti-cancer approach.

Recent studies have provided evidence for the existence of at least one tissue-specific angiogenic factor (EG-VEGF) expressed by a steroidogenic gland⁹⁵. Highly vascularized endocrine malignancies such as Leydig cell tumours express particularly high levels of EG-VEGF mRNA⁹⁶, raising the possibility that EG-VEGF inhibition may have a role in the treatment of such tumours. It remains to be established whether other tissue-specific angiogenic pathways exist.

Furthermore, an emerging concept in solid-tumour biology is that of tumour 'stem' cells. It has been argued that this minor self-renewing subpopulation is absolutely crucial to progressive and sustained cancer growth, as opposed to the bulk non-stem-cell population^{97,98}. This may explain why rapid and significant tumour shrinkage induced by various anti-cancer drugs such as cytotoxic chemotherapy may have little or no impact on prolonging survival: the 'wrong' population is being eliminated. By contrast, antiangiogenic drugs, which generally do not cause rapid tumour shrinkage, if any shrinkage at all, may nevertheless have a survival benefit if they block the stem-cell population from inducing or exploiting angiogenesis necessary for their growth. Indeed, there is already a hint that the stem-cell subpopulation in breast cancer may be more adept at promoting angiogenesis than its non-stem-cell counterpart⁹⁹.

- Ferrara, N. VEGF and the quest for tumour angiogenesis factors. *Nature Rev. Cancer* **2**, 795–803 (2002).
- Ide, A. G., Baker, N. H. & Warren, S. L. Vascularization of the Brown Pearce rabbit epithelioma transplant as seen in the transparent ear chamber. *Am. J. Roentgenol.* **42**, 891–899 (1939).
- Algire, G. H., Chalkley, H. W., Legallais, F. Y. & Park, H. D. Vascular reactions of normal and malignant tissues *in vivo*. I. Vascular reactions of mice to wounds and to normal and neoplastic transplants. *J. Natl. Cancer Inst.* **6**, 73–85 (1945).
- Folkman, J. Tumor angiogenesis: therapeutic implications. *N. Engl. J. Med.* **285**, 1182–1186 (1971).
- Hurwitz, H. et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N. Engl. J. Med.* **350**, 2335–2342 (2004).
- Gragoudas, E. S., Adamis, A. P., Cunningham, E. T. Jr, Feinsod, M. & Guyer, D. R. Pegaptanib for neovascular age-related macular degeneration. *N. Engl. J. Med.* **351**, 2805–2816 (2004).
- Kerbel, R. & Folkman, J. Clinical translation of angiogenesis inhibitors. *Nature Rev. Cancer* **2**, 727–739 (2002).
- Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* **9**, 669–676 (2003).
- Yancopoulos, G. D. et al. Vascular-specific growth factors and blood vessel formation. *Nature* **407**, 242–248 (2000).
- Ferrara, N. Vascular endothelial growth factor: basic science and clinical progress. *Endocr. Rev.* **25**, 581–611 (2004).
- Hattori, K. et al. Placental growth factor reconstitutes hematopoiesis by recruiting VEGFR1+ stem cells from bone-marrow microenvironment. *Nature Med.* **8**, 841–849 (2002).
- Gerber, H.-P. et al. Vascular endothelial growth factor regulates haematopoietic stem cell survival by an internal autocrine loop mechanism. *Nature* **417**, 954–958 (2002).
- Luttun, A. et al. Revascularization of ischemic tissues by PlGF treatment, and inhibition of tumor angiogenesis, arthritis and atherosclerosis by anti-Flt1. *Nature Med.* **8**, 831–840 (2002).
- Hiratsuka, S. et al. MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis. *Cancer Cell* **2**, 289–300 (2002).
- LeCouter, J. et al. Angiogenesis-independent endothelial protection of liver: role of VEGFR-1. *Science* **299**, 890–893 (2003).
- Wey, J. S. et al. Vascular endothelial growth factor receptor-1 promotes migration and invasion in pancreatic carcinoma cell lines. *Cancer* **104**, 427–438 (2005).
- Safran, M. & Kaelin, W. J. Jr. HIF hydroxylation and the mammalian oxygen-sensing pathway. *J. Clin. Invest.* **111**, 779–783 (2003).
- Dvorak, H. F. Vascular permeability factor/vascular endothelial growth factor: a critical cytokine in tumor angiogenesis and a potential target for diagnosis and therapy. *J. Clin. Oncol.* **20**, 4368–4380 (2002).
- Lonsler, R. R. et al. von Hippel-Lindau disease. *Lancet* **361**, 2059–2067 (2003).
- Yang, J. C. et al. A randomized trial of bevacizumab, an anti-VEGF antibody, for metastatic renal cancer. *N. Engl. J. Med.* **349**, 427–434 (2003).
- Rak, J. et al. Mutant ras oncogenes upregulate VEGF/VPF expression: implications for induction and inhibition of tumor angiogenesis. *Cancer Res.* **55**, 4575–4580 (1995).
- Brugarolas, J. & Kaelin, W. G. Jr. Dysregulation of HIF and VEGF is a unifying feature of the familial hamartoma syndromes. *Cancer Cell* **6**, 7–10 (2004).
- Kim, K. J. et al. Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumor growth *in vivo*. *Nature* **362**, 841–844 (1993).
- Gerber, H. P. & Ferrara, N. Pharmacology and pharmacodynamics of bevacizumab as monotherapy or in combination with cytotoxic therapy in preclinical studies. *Cancer Res.* **65**, 671–680 (2005).
- Millauer, B., Shawver, L. K., Plate, K. H., Risau, W. & Ullrich, A. Glioblastoma growth inhibited *in vivo* by a dominant-negative Flk-1 mutant. *Nature* **367**, 576–579 (1994).
- Prewett, M. et al. Antivascular endothelial growth factor receptor (fetal liver kinase 1) monoclonal antibody inhibits tumor angiogenesis. *Cancer Res.* **59**, 5209–5218 (1999).
- Wood, J. M. et al. PTK787/ZK 222584, a novel and potent inhibitor of vascular endothelial growth factor receptor tyrosine kinases, impairs vascular endothelial growth factor-induced responses and tumor growth after oral administration. *Cancer Res.* **60**, 2178–2189 (2000).
- Gerber, H. P., Kowalski, J., Sherman, D., Eberhard, D. A. & Ferrara, N. Complete inhibition of rhabdomyosarcoma xenograft growth and neovascularization requires blockade of both tumor and host vascular endothelial growth factor. *Cancer Res.* **60**, 6253–6258 (2000).
- Holash, J. et al. VEGF-Trap: a VEGF blocker with potent antitumor effects. *Proc. Natl. Acad. Sci. USA* **99**, 11393–11398 (2002).
- Inoue, M., Hager, J. H., Ferrara, N., Gerber, H. P. & Hanahan, D. VEGF-A has a critical, non-redundant role in angiogenic switching and pancreatic β cell carcinogenesis. *Cancer Cell* **1**, 193–202 (2002).
- Lindhal, P., Johansson, B. E., Leveen, P. & Betsholtz, C. Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* **277**, 242–245 (1997).
- Bergers, G., Song, S., Meyer-Morse, N., Bergsland, E. & Hanahan, D. Benefits of targeting both pericytes and endothelial cells in the tumor vasculature with kinase inhibitors. *J. Clin. Invest.* **111**, 1287–1295 (2003).
- Dong, J. et al. VEGF-null cells require PDGFR alpha signaling-mediated stromal fibroblast recruitment for tumorigenesis. *EMBO J.* **23**, 2800–2810 (2004).
- Maisonpierre, P. C. et al. Angiopoietin-2, a natural antagonist for Tie-2 that disrupts *in vivo* angiogenesis. *Science* **277**, 55–60 (1997).
- Oliner, J. et al. Suppression of angiogenesis and tumor growth by selective inhibition of angiopoietin-2. *Cancer Cell* **6**, 507–516 (2004).
- Carmeliet, P. & Tessier-Lavigne, M. Common mechanisms of nerve and blood vessel wiring. *Nature* **436**, 193–200 (2005).
- Klagsbrun, M. & Eichmann, A. A role for axon guidance receptors and ligands in blood vessel development and tumor angiogenesis. *Cytokine Growth Factor Rev.* **16**, 535–548 (2005).
- Soker, S., Takashima, S., Miao, H. Q., Neufeld, G. & Klagsbrun, M. Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell* **92**, 735–745 (1998).
- Pasquale, E. B. Eph receptor signalling casts a wide net on cell behaviour. *Nature Rev. Mol. Cell. Biol.* **6**, 462–475 (2005).
- Pandey, A., Shao, H., Marks, R. M., Polverini, P. J. & Dixit, V. M. Role of B61, the ligand for the Eck receptor tyrosine kinase, in TNF-alpha-induced angiogenesis. *Science* **268**, 567–569 (1995).
- Martiny-Baron, G. et al. Inhibition of tumor growth and angiogenesis by soluble EphA2. *Neoplasia* **6**, 248–257 (2004).
- Dobrzanski, P. et al. Antiangiogenic and antitumor efficacy of EphA2 receptor antagonist. *Cancer Res.* **64**, 910–919 (2004).
- Wang, B. et al. Induction of tumor angiogenesis by Slit-Robo signaling and inhibition of cancer growth by blocking Robo activity. *Cancer Cell* **4**, 19–29 (2003).
- DiPietro, L. A. Thrombospondin as a regulator of angiogenesis. In *Regulation of Angiogenesis* (eds Rosen, E. & Goldberg, I. D.) 295–314 (Springer, Berlin, 1997).
- O'Reilly, M. S. et al. Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* **88**, 277–285 (1997).
- Sund, M. et al. Function of endogenous inhibitors of angiogenesis as endothelium-specific tumor suppressors. *Proc. Natl. Acad. Sci. USA* **102**, 2934–2939 (2005).
- Pike, S. E. et al. Vasostatin, a calreticulin fragment, inhibits angiogenesis and suppresses tumor growth. *J. Exp. Med.* **188**, 2349–2356 (1998).
- Watanabe, K. et al. Vasohibin as an endothelium-derived negative feedback regulator of angiogenesis. *J. Clin. Invest.* **114**, 898–907 (2004).
- Nyberg, P., Xie, L. & Kalluri, R. Endogenous inhibitors of angiogenesis. *Cancer Res.* **65**, 3967–3979 (2005).
- Asahara, T. et al. Isolation of putative progenitor endothelial cells for angiogenesis. *Science* **275**, 964–967 (1997).
- Takahashi, T. et al. Ischemia- and cytokine-induced mobilization of bone marrow-derived endothelial progenitor cells for neovascularization. *Nature Med.* **5**, 434–438 (1999).
- Lyden, D. et al. Impaired recruitment of bone-marrow-derived endothelial and hematopoietic precursor cells blocks tumor angiogenesis and growth. *Nature Med.* **7**, 1194–1201 (2001).
- Heissig, B. et al. Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of kit-ligand. *Cell* **109**, 625–637 (2002).
- Ruzinova, M. B. et al. Effect of angiogenesis inhibition by Id loss and the contribution of bone-marrow-derived endothelial cells in spontaneous murine tumors. *Cancer Cell* **4**, 277–289 (2003).
- De Palma, M., Veneri, M. A., Roca, C. & Naldini, L. Targeting exogenous genes to tumor angiogenesis by transplantation of genetically modified hematopoietic stem cells. *Nature Med.* **9**, 789–795 (2003).
- Peters, B. A. et al. Contribution of bone marrow-derived endothelial cells to human tumor vasculature. *Nature Med.* **11**, 261–262 (2005).
- Shaked, Y. et al. Genetic heterogeneity of the vasculogenic phenotype parallels angiogenesis: Implications for cellular surrogate marker analysis of antiangiogenesis. *Cancer Cell* **7**, 101–111 (2005).
- Gasparini, G., Longo, R., Fanelli, M. & Teicher, B. A. Combination of antiangiogenic therapy with other anticancer therapies: results, challenges, and open questions. *J. Clin. Oncol.* **23**, 1295–1311 (2005).
- Browder, T. et al. Antiangiogenic scheduling of chemotherapy improves efficacy against experimental drug-resistant cancer. *Cancer Res.* **60**, 1878–1886 (2000).
- Klement, G. et al. Continuous low-dose therapy with vinblastine and VEGF receptor-2 antibody induces sustained tumor regression without overt toxicity. *J. Clin. Invest.* **105**, R15–R24 (2000).
- Sweeney, C. J. et al. The antiangiogenic property of docetaxel is synergistic with a recombinant human monoclonal antibody against vascular endothelial growth factor or 2-methoxyestradiol but antagonized by endothelial growth factors. *Cancer Res.* **61**, 3369–3372 (2001).
- Bertolini, F. et al. Maximum tolerable dose and low-dose metronomic chemotherapy have opposite effects on the mobilization and viability of circulating endothelial progenitor cells. *Cancer Res.* **63**, 4342–4346 (2003).

63. Kerbel, R. S. & Kamen, B. A. The anti-angiogenic basis of metronomic chemotherapy. *Nature Rev. Cancer* **4**, 423–436 (2004).
64. Kim, J. J. & Tannock, I. F. Repopulation of cancer cells during therapy: an important cause of treatment failure. *Nature Rev. Cancer* **5**, 516–525 (2005).
65. Jain, R. K. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* **307**, 58–62 (2005).
66. Winkler, F. *et al.* Kinetics of vascular normalization by VEGFR2 blockade governs brain tumor response to radiation: role of oxygenation, angiopoietin-1, and matrix metalloproteinases. *Cancer Cell* **6**, 553–563 (2004).
67. Willett, C. G. *et al.* Direct evidence that the VEGF-specific antibody bevacizumab has antivasculature effects in human rectal cancer. *Nature Med.* **10**, 145–147 (2004).
68. Siemann, D. W., Chaplin, D. J. & Horsman, M. R. Vascular-targeting therapies for treatment of malignant disease. *Cancer* **100**, 2491–2499 (2004).
69. Ferrara, N., Hillan, K. J., Gerber, H. P. & Novotny, W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nature Rev. Drug Discov.* **3**, 391–400 (2004).
70. Miller, K. D. *et al.* Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. *J. Clin. Oncol.* **23**, 792–799 (2005).
71. Johnson, D. H. *et al.* Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. *J. Clin. Oncol.* **22**, 2184–2191 (2004).
72. Smith, J. K., Mamoon, N. M. & Duhe, R. J. Emerging roles of targeted small molecule protein-tyrosine kinase inhibitors in cancer therapy. *Oncol. Res.* **14**, 175–225 (2004).
73. Herbst, R. S. *et al.* Phase I study of recombinant human endostatin in patients with advanced solid tumors. *J. Clin. Oncol.* **20**, 3792–3803 (2002).
74. Garner, A. Vascular diseases. In *Pathobiology of Ocular Disease* (eds Garner, A. & Klintworth, G. K.) 1625–1710 (Marcel Dekker, New York, 1994).
75. Ruckman, J. *et al.* 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *J. Biol. Chem.* **273**, 20556–20567 (1998).
76. Chen, Y. *et al.* Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured Fab in complex with antigen. *J. Mol. Biol.* **293**, 865–881 (1999).
77. Kerbel, R. S. *et al.* Possible mechanisms of acquired resistance to anti-angiogenic drugs: implications for the use of combination therapy approaches. *Cancer Metastasis Rev.* **20**, 79–86 (2001).
78. Sweeney, C. J., Miller, K. D. & Sledge, G. W. Jr. Resistance in the anti-angiogenic era: naysaying or a word of caution? *Trends Mol. Med.* **9**, 24–29 (2003).
79. Yu, J. L., Rak, J. W., Coomber, B. L., Hicklin, D. J. & Kerbel, R. S. Effect of p53 status on tumor response to antiangiogenic therapy. *Science* **295**, 1526–1528 (2002).
80. Glade Bender, J., Cooney, E. M., Kandel, J. J. & Yamashiro, D. J. Vascular remodeling and clinical resistance to antiangiogenic cancer therapy. *Drug Resist. Updat.* **7**, 289–300 (2004).
81. Orimo, A. *et al.* Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**, 335–348 (2005).
82. Lewis, C. & Murdoch, C. Macrophage responses to hypoxia: implications for tumor progression and anti-cancer therapies. *Am. J. Pathol.* **167**, 627–635 (2005).
83. Guo, W. & Giancotti, F. G. Integrin signalling during tumour progression. *Nature Rev. Mol. Cell. Biol.* **5**, 816–826 (2004).
84. Hida, K. *et al.* Tumor-associated endothelial cells with cytogenetic abnormalities. *Cancer Res.* **64**, 8249–8255 (2004).
85. Streubel, B. *et al.* Lymphoma-specific genetic aberrations in microvascular endothelial cells in B-cell lymphomas. *N. Engl. J. Med.* **351**, 250–259 (2004).
86. Ferrara, N. & Alitalo, K. Clinical applications of angiogenic growth factors and their inhibitors. *Nature Med.* **5**, 1359–1364 (1999).
87. Simons, M. Angiogenesis: where do we stand now? *Circulation* **111**, 1556–1566 (2005).
88. Lederman, R. J. *et al.* Therapeutic angiogenesis with recombinant fibroblast growth factor-2 for intermittent claudication (the TRAFFIC study): a randomised trial. *Lancet* **359**, 2053–2058 (2002).
89. Henry, T. D. *et al.* The VIVA trial: Vascular endothelial growth factor in ischemia for vascular angiogenesis. *Circulation* **107**, 1359–1365 (2003).
90. Heinl-Green, A. *et al.* The efficacy of a 'master switch gene' HIF-1 α in a porcine model of chronic myocardial ischaemia. *Eur. Heart J.* **26**, 1327–1332 (2005).
91. Rafii, S. & Lyden, D. Therapeutic stem and progenitor cell transplantation for organ vascularization and regeneration. *Nature Med.* **9**, 702–712 (2003).
92. Tateishi-Yuyama, E. *et al.* Therapeutic angiogenesis for patients with limb ischaemia by autologous transplantation of bone-marrow cells: a pilot study and a randomised controlled trial. *Lancet* **360**, 427–435 (2002).
93. Romond, E. H. *et al.* Trastuzumab plus adjuvant chemotherapy for operable HER-2-positive breast cancer. *N. Engl. J. Med.* **353**, 1673–1684 (2005).
94. Morgan, B. *et al.* Dynamic contrast-enhanced magnetic resonance imaging as a biomarker for the pharmacological response of PTK787/ZK 222584, an inhibitor of the vascular endothelial growth factor receptor tyrosine kinases, in patients with advanced colorectal cancer and liver metastases: results from two phase I studies. *J. Clin. Oncol.* **21**, 3955–3964 (2003).
95. LeCouter, J. *et al.* Identification of an angiogenic mitogen selective for endocrine gland endothelium. *Nature* **412**, 877–884 (2001).
96. Samson, M. *et al.* Human endocrine gland-derived vascular endothelial growth factor: expression early in development and in Leydig cell tumors suggests roles in normal and pathological testis angiogenesis. *J. Clin. Endocrinol. Metab.* **89**, 4078–4088 (2004).
97. Webb, T. Work on breast cancer stem cells raises questions about treatment strategies. *J. Natl. Cancer Inst.* **95**, 774–775 (2003).
98. Jones, R. J., Matsui, W. H. & Smith, B. D. Cancer stem cells: are we missing the target? *J. Natl. Cancer Inst.* **96**, 583–585 (2004).
99. Ponti, D. *et al.* Isolation and *in vitro* propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res.* **65**, 5506–5511 (2005).

Acknowledgements R.S.K. was supported by grants from the Canadian Institutes for Health Research and the National Institute of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence should be addressed to N.F. (nf@gene.com).

Probing ion-channel pores one proton at a time

Gisela D. Cymes¹, Ying Ni¹ & Claudio Grosman¹

Although membrane proteins often rely on ionizable residues for structure and function, their ionization states under physiological conditions largely elude experimental estimation. To gain insight into the effect of the local microenvironment on the proton affinity of ionizable residues, we have engineered individual lysines, histidines and arginines along the α -helical lining of the transmembrane pore of the nicotinic acetylcholine receptor. We can detect individual proton binding-unbinding reactions electrophysiologically at the level of a single proton on a single side chain as brief blocking-unblocking events of the passing cation current. Kinetic analysis of these fluctuations yields the position-dependent rates of proton transfer, from which the corresponding pK_a values and shifts in pK_a can be calculated. Here we present a self-consistent, residue-by-residue description of the microenvironment around the pore-lining transmembrane α -helices (M2) in the open-channel conformation, in terms of the excess free energy that is required to keep the engineered basic side chains protonated relative to bulk water. A comparison with closed-channel data leads us to propose that the rotation of M2, which is frequently invoked as a hallmark of the gating mechanism of Cys-loop receptors, is minimal, if any.

Electrostatic interactions have an important role in diverse aspects of protein structure and function^{1–4}, ranging from enzyme catalysis, ligand binding and the fine-tuning of redox potentials, to the stability of folded proteins and the translocon-mediated integration of transmembrane segments into the endoplasmic reticulum. These interactions often involve full charges on the side chains of ionizable amino acids that arise from the association and dissociation of protons to and from these acid–base groups in the protein. In ion channels, charges on ionizable side chains control single-channel conductance^{5,6}, ion selectivity^{7–9}, open-channel block^{10,11}, gating¹² and voltage sensing^{13,14}. Knowledge of the protonation state of these residues is required for a thorough understanding of these electrostatically controlled phenomena in quantitative detail, but the pK_a values of ionizable side chains in ion channels, as well as in most other large membrane proteins, continue to elude experimental determination. In addition, because the probability of these side chains bearing a charge at a given pH is a function of the electrostatic properties of the surrounding microenvironment, the well-known pK_a values of these side chains in bulk solution cannot be used to predict protonation states in any but the most trivial situations (that is, when the protonatable group is surface-exposed). Furthermore, it is well recognized that the uncertainty associated with theoretically calculated pK_a values increases with the extent to which the side chain in question is buried^{15,16}, a problem that, we gather, might be even more acute in membrane proteins owing to the paucity of experimentally estimated benchmark values.

Single-side-chain acid-base titrations

We reasoned that, owing to their unique properties, ion channels provide an extraordinary opportunity to investigate the effect of the protein microenvironment on the (time-averaged) pK_a values of ionizable residues located in pore domains. In principle, individual protonation–deprotonation events, which would cause the net charge of the pore to oscillate by one unit, should be evident in a

single-channel recording as current fluctuations between two levels: one corresponding to the ‘neutral’ pore; the other corresponding to the pore carrying the extra unit charge. For example, for a single basic residue (lysine, arginine or histidine) engineered in a cation-selective pore, the ‘deprotonated channel’ is expected to have roughly the same conductance as the wild-type protein, whereas the channel with the protonated (positively charged) residue is expected to conduct cations at a lower rate. Protonation–deprotonation events occurring while the channel is open should thus be manifest as oscillations of the current between two discrete values. The ratio between these two values would be a measure of the proximity of the protonatable group to the long axis of the pore, and the kinetics of the current fluctuations would reflect directly the rates of proton transfer and the pK_a of the side chain.

The actual detection of such individual proton-transfer events is, of course, not a foregone conclusion. The rates of proton binding and unbinding might be too fast, for example, exceeding the temporal resolution of existing single-channel recording techniques (a few microseconds under optimal conditions¹⁷). Alternatively, the extent to which a charge added to the pore affects the conductance of the channel might be too small for any proton-transfer event to be detected as a discrete step change in the current. Even more elementarily, because buried charges can reduce the stability of folded proteins¹⁸, engineered ionizable side chains could prevent the correct folding and insertion of the protein into the membrane.

The literature on ion channels, however, contains at least two examples for which this phenomenon has indeed been recorded at the single-channel level, namely L-type Ca^{2+} channels¹⁹ and cyclic-nucleotide-gated (CNG) channels²⁰. In these two examples, the relevant ionizable groups corresponded to the naturally occurring rings of four acidic residues present in the P-loop region of these ion channels. Although proton-transfer rates were measurable (they were exceedingly fast in aqueous solutions but were slowed down to measurable values by changing the solvent to deuterium oxide),

¹Department of Molecular and Integrative Physiology, Center for Biophysics and Computational Biology, and Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

the protonation–deprotonation of the individual carboxylate–carboxylic groups could not be identified. Rather, the four acidic side chains seemed to interact to form a single (Ca^{2+} channel) or two (CNG channel) proton-binding sites.

Engineering ionizable side chains

As a step towards understanding the effect of the microenvironment on the pK_a values of membrane protein residues, we systematically engineered protonatable side chains into the pore domain of the adult-type muscle nicotinic acetylcholine receptor (AChR), a member of the Cys-loop receptor superfamily of neurotransmitter-gated ion channels (Fig. 1a and Supplementary Fig. 1). This receptor offers many technical advantages for this type of study, including a large conductance ($\sim 84\text{ pS}$), the presence of a single copy of three of its four different subunits and an invariant stoichiometry, $(\alpha 1)_2\beta 1\delta \epsilon$.

In this superfamily of pentameric receptor channels, the pore is formed by the lateral association of five α -helical transmembrane segments, termed M2 (ref. 21), the rotation of which has been proposed to underlie the closed to open transition^{22–24}. Acidic and basic residues flank the M2 segments of all the members of the Cys-loop superfamily where, in addition to having topogenic effects that are common to all transmembrane segments, they are crucial determinants of charge selectivity, conductance and rectification^{5,9,25}.

In marked contrast, protonatable residues are very rare in the membrane-spanning portion of M2 in these channels and are completely absent in nicotinic receptors. Although protonation states for some of these naturally occurring residues have been suggested on the basis of electrostatic model calculations²⁶, the need for experimental approaches is clear.

First, we introduced single lysines, one at a time, into a continuous stretch of 32 residues of the δ -subunit of the AChR, starting with Pro 250 (position $-7'$; Fig. 1a) at the amino-terminal end of the M1–M2 loop, and ending with Thr 281 (position $25'$) close to the carboxy-terminal end of the extracellular portion of the M2 helix. The naturally occurring lysines occupying positions $0'$ and $20'$ (Fig. 1a) were mutated to alanine, whereas the aspartate at position $-5'$, the glutamate at $-1'$ and the arginine at $21'$ were mutated to both alanine and lysine. All constructs were successfully expressed in HEK-293 cells and the current-blocking effect of single protonated lysines was indeed observed and analysed in detail.

Figure 1b–d summarizes the effect of the $\epsilon\text{NH}_2/\text{NH}_3^+$ group of lysine on the single-channel conductance of the AChR. Insofar as the extent of channel block (Fig. 1d) can be taken as a rough measure of distance between the engineered charges and the long axis of the pore, this plot can be interpreted in structural terms. Thus, an uninterrupted α -helical pattern is apparent between positions $1'$

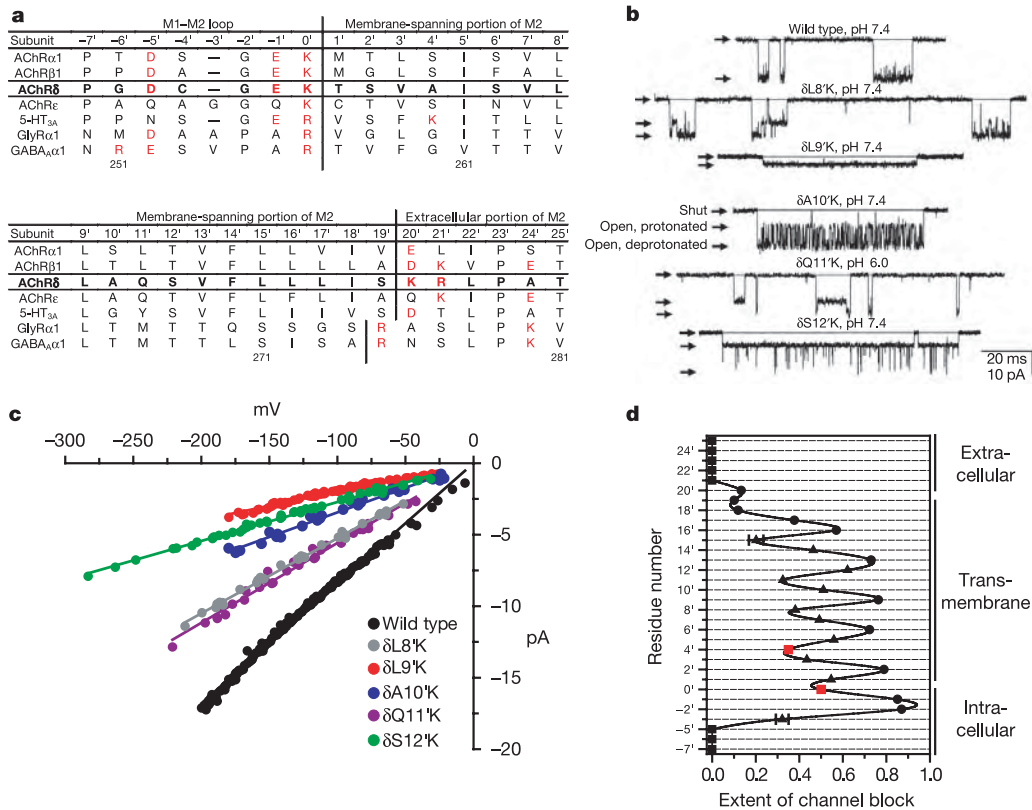


Figure 1 | Protonation of lysines causes partial channel block. **a**, Sequence alignment of the portion of M2 studied here. The four adult mouse muscle AChR subunits ($\alpha 1$, $\beta 1$, δ and ϵ) and one representative each of the serotonin (5-HT $_3$), GABA and glycine (GlyR) receptor channel subunits are compared. Protonatable residues are indicated in red. AChR δ -subunit residues are indicated in bold. The prime numbering system is indicated at the top, and the specific residue numbers of the mouse δ -subunit are indicated at the bottom. **b**, Examples of single-channel inward currents ($V \approx -200\text{ mV}$, $1\text{ }\mu\text{M}$ ACh) recorded from HEK-293 cells expressing the indicated constructs. Openings are downward deflections. Arrows indicate the shut (zero current) and the two open current levels. The shut level of each trace is also indicated by a horizontal broken line. Display $f_c = 6\text{ kHz}$. **c**, Current–voltage (I – V) relationships for the five mutants in **b** and the wild-type AChR. For clarity, only the I – V curve corresponding to the

blocked open state is shown for each mutant. **d**, Extent of channel block for each lysine construct calculated as the difference in conductance between the main level and the sublevel, normalized by the conductance of the main level ($\Delta\text{conductance}/\text{main-level conductance}$). For positions $0'$ and $4'$, the extent-of-block values (red squares) are predictions based on the values observed at neighbouring positions (see text and Supplementary Methods). The unbroken line is a cubic-spline interpolation. Proposed membrane boundaries²³ are tentative: our data would not be inconsistent with these being displaced by roughly one turn towards the intracellular. Note that even positions that would be on the ‘back’ of the α -helix (such as $8'$, $11'$ and $15'$) exert a considerable electrostatic effect on the cation current. For most positions, the horizontal error bars (standard errors) are smaller than the experimental points. The data suggest that the lumen of the pore is to the right of the plot.

and 17' (perhaps even between $-2'$ and $18'$), with the narrowest constriction being formed by positions $-2'$ and $-1'$, and with the electrostatic effect of positive unit charges on the permeating cations decreasing steeply outside the presumed membrane-spanning region. Figure 1d also suggests that there is a particular rotational angle for the M2 α -helix in the open-channel conformation. Positions $-2'$, $-1'$, $2'$, $6'$, $9'$, $13'$ and $16'$ seem to form the lumen-facing side of the α -helix in the open state, as lysine substitutions have their greatest blocking effect when engineered at these locations. The orientation of the other residues can be similarly inferred from this plot. The secondary structure of M2 beyond the stretch between approximately $-4'$ and $18'$ could not be ascertained because the blocking effect of engineered charges becomes too small.

On the basis of whether protonation–deprotonation events of the introduced single lysines could be observed as fluctuations of the open-channel current between two levels of different conductance (Figs 1b and 2), all probed positions were classified in three groups: (1) the open-channel current seems to be fixed at a conductance level lower than that of wild type at all pH values tested (6.0, 7.4 and 9.0), with excursions to a higher-conductance level becoming slightly more apparent, but remaining marginally quantifiable, as the pH is raised; (2) the open-channel current fluctuates between a wild-type-like ('main') and a lower ('sub') conductance level in a pH-dependent manner, with the occupancy of the main level increasing as the pH increases; and (3) the open-channel current seems to remain fixed at, or very close to, the wild-type level at all pH values tested (6.0, 7.4 and 9.0).

Positions $-2'$, $-1'$, $2'$, $6'$, $9'$, $13'$, $16'$, $17'$, $18'$, $19'$ and $20'$ fall into the first group (Fig. 1d, circles). The reduced conductance and the failure to detect open-channel current fluctuations are consistent with the introduced lysines being predominantly protonated, whereas the pH insensitivity in the pH range 6.0–9.0 (Fig. 3a) suggests that the pK_a values are as high or perhaps even higher than that of the lysine side chain in bulk water (~ 10.4). Hence,

precise pK_a values could not be estimated at these positions using the ϵNH_2 group of lysine as a probe.

Positions $-4'$, $1'$, $3'$, $5'$, $7'$, $8'$, $10'$, $11'$, $12'$, $14'$, and $15'$ fall into the second group (Fig. 1d, triangles). The observed pH dependence of the main-level to sublevel current fluctuations (Fig. 3c) is compelling evidence that this phenomenon results from the protonation–deprotonation of individual lysine side chains. The fact that proton-transfer events involving lysines ($pK_{a,bulk} \approx 10.4$) can be observed indicates that the local microenvironment around these positions lowers the pK_a of the ϵNH_2 group, bringing it closer to the 6.0–9.0 pH range used in the experiments. From the estimated rates of proton association and dissociation, we calculated the side chain pK_a values at all of these locations (Fig. 2 and Table 1; see also Supplementary Figs 2 and 3c, d, and Supplementary Methods) with the exception of positions $-4'$, $5'$ and $15'$, for which the current fluctuations were difficult to resolve.

Positions $-7'$, $-6'$, $-5'$, $0'$, $4'$ and the continuous stretch from $21'$ to $25'$ fall into the third group (Fig. 1d, squares). A wild-type-like conductance and the absence of open-channel current fluctuations are consistent with a very acidic pK_a for the lysine side chain, such that the probability of being protonated, and hence of blocking the cation current, is negligibly small. However, this behaviour is also consistent with an orientation (such as a large distance from the long axis of the pore) and/or a local microenvironment (for example, water with bulk-like properties) that markedly screen the effect of the ϵNH_3^+ charge on the passing current, such that protonation–deprotonation events would go undetected even if they occurred. These two different situations—namely, a highly downshifted pK_a or an attenuated electrostatic effect—cannot be distinguished easily on the basis of our data. We favour the notion that the latter might be true for positions $-7'$, $-6'$ and $-5'$ on the distal end of the cytoplasmic M1–M2 linker, and for the stretch between $21'$ and $25'$ located on the portion of M2 that protrudes above the membrane into the extracellular space²³.

The situation of positions $0'$ and $4'$ is, however, less obvious. On

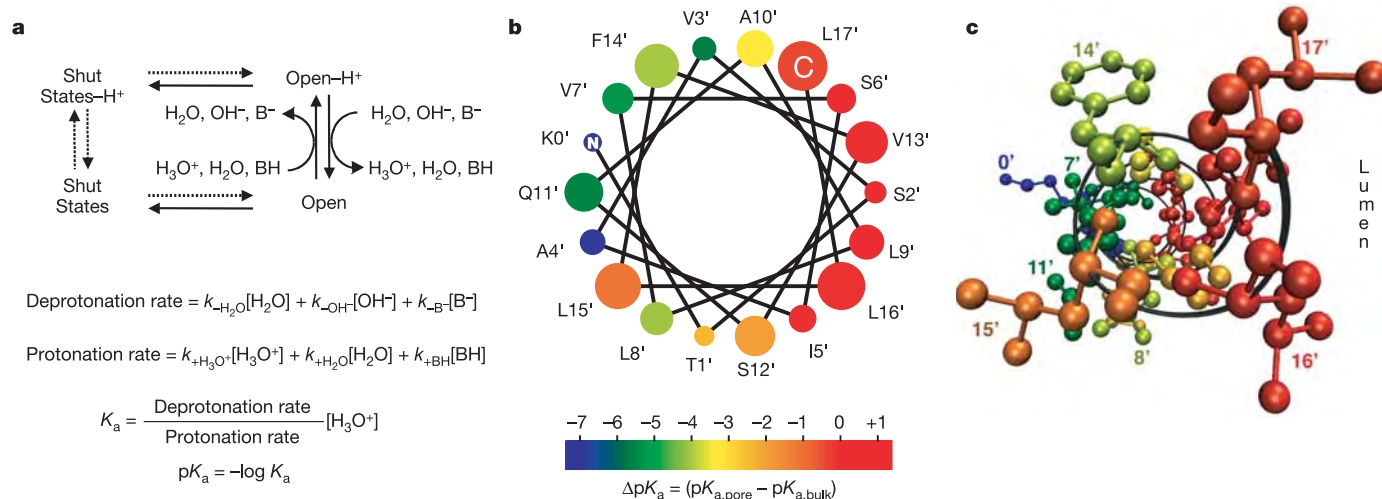


Figure 2 | Experimental estimates of pK_a values. **a**, Conceptual framework of this work: a simple thermodynamic cycle. The channel interconverts among closed, desensitized (collectively referred to as 'shut states') and open conformations with or without an extra proton bound to the pore domain. The association and dissociation of a single proton to the open state (but not to the shut states) is manifest as a discrete change in the rate of ion flow. Kinetic analysis of these fluctuations, and of the open to shut transitions, yields the rates of protonation and deprotonation of the open channel; these rates, along with the pH of the solution, are then combined to calculate the pK_a of the engineered residue. The three proton donors and three proton acceptors present in the solutions are indicated. BH and B^- denote the protonated and deprotonated forms of the H^+ buffer, respectively. Note that

the kinetics of both proton transfer and channel shutting affect the duration of sojourns in the sublevel and main level (unbroken arrows). **b**, ΔpK_a values mapped onto an ideal α -helical wheel representation of M2 of the δ -subunit. The size of the symbols increases towards the extracellular end. Wild-type residues and the C and N termini are indicated. The lumen of the pore would be to the right of the plot. The unique properties of position $15'$ are discussed in the Supplementary Discussion. **c**, ΔpK_a values mapped onto a ball-and-stick representation of the $0'$ – $17'$ positions of the δ -subunit M2 α -helix, generated using atomic coordinates in the PDB file 1OED²³. The colour code is the same as in **b**. The backbone is indicated in ribbon representation. The molecular image was made with VMD⁴³.

the one hand, if we assume that the periodicity of the blocking effect (Fig. 1d) should be maintained without interruption throughout the α -helical segment, then a highly depressed pK_a value can explain why these two lysines do not block the current. If this were the case, then the red square symbols in Fig. 1d show the rough extent of channel block that lysines at 0' and 4' would cause if they were protonated. On the other hand, we cannot dismiss the possibility that, owing to the long (~ 8 Å) and flexible side chain of lysine, the 0' and 4' ϵNH_2 groups reach into the bulk-like intracellular water and become protonated, in a process akin to the 'snorkelling' of lysines and arginines located at the ends of lipid-embedded α -helices²⁷. Screening of the ϵNH_3^+ charge by the highly polarizable environment would thus account for the lack of observable block. However, the finding that lysines at positions 1' and 3' do not seem to 'snorkel' (that is, their extent-of-block values are not zero; Fig. 1d), even when present in rather hydrophobic microenvironments (Fig. 2b, c and Table 1), reduces the likelihood of the 'snorkelling' notion. This seems particularly true for position 4', which is located more deeply in the membrane. Clearly, more experiments and electrostatic calculations are needed to settle this point.

To extend our pK_a measurements to locations where the introduced lysine side chains seemed to be permanently protonated, we engineered histidines at positions -2', -1', 2', 6', 9', 13', 16' and 17' (positions 18', 19' and 20' were not investigated further because the small extent of block (Fig. 1d) renders the analysis unreliable). Because the 'macroscopic' pK_a of the imidazole ring of histidine in bulk solution is ~ 6.4 , the imidazole to imidazolium⁺ interconversion should be detected at these 'bulk-like' positions in the pH range 6.0–9.0. Indeed, pH-dependent main-level to sublevel fluctuations (Fig. 3b) were recorded for all of these mutants, and the corresponding pK_a values were calculated from the kinetics of proton

transfer (Fig. 2 and Table 1; see also Supplementary Figs 2 and 3a, b, and Supplementary Methods). As qualitatively expected from the corresponding lysine substitutions, the imidazole pK_a values are not depressed, consistent with these positions facing the aqueous lumen of the permeation pathway. In fact, at most of these positions, the pK_a values are higher than the bulk value of 6.4, which indicates that it is easier to charge a histidine in this location within the pore than in bulk water. Provided that this is also the case for the ϵNH_2 group of lysine, this explains our initially puzzling observation that the subconductance levels of lysine constructs at these positions were oblivious to pH, even (as was the case for the 6' and 9' substitutions) when exposed to pH values as high as 10.5.

We also introduced histidines at -4', 5' and 15', three positions where lysine-induced current fluctuations were observed but were difficult to quantify reliably. The corresponding histidine pK_a values and rates of proton transfer are shown in Table 1. We also engineered a histidine at position 12'. The local microenvironment around this position lowers the affinity of the lysine side chain for protons by a factor of ~ 30 relative to that in bulk water ($\Delta pK_a \approx 8.9 - 10.4 = -1.5$; Fig. 2 and Table 1). Thus, to the extent that the side chains of histidine and lysine probe similar microenvironments, and that their pK_a values have similar sensitivities to them, a histidine at 12' is expected to be mostly deprotonated ($pK_a \approx 6.4 - 1.5 = 4.9$) even at pH 6.0. Indeed, our single-channel recordings clearly confirmed this prediction (Fig. 3d), and we generalize this result to suggest that histidines engineered at positions that lower the pK_a of lysine to the same or a greater degree than at position 12' ($\Delta pK_a < -1.5$ units; see colour code in Fig. 2b, c) would be largely neutral, even at pH 6.0.

To gain further insight into the unique microenvironment around positions 0' and 4', we engineered arginines at these positions. As for the 0' and 4' lysine constructs, arginine substitutions at these two positions did not reduce the single-channel conductance. This result confirms the notion that charging the pore at these two positions might be energetically costly (see above) such that not even the guanidine group of arginine, which has a bulk pK_a of ~ 12.0 , can bind a proton at pH 6.0. We conclude that the pK_a shift would be at least -7 units (Fig. 2 and Table 1). For comparison, we also engineered an arginine at 11' (on the same face of the α -helix as 0' and 4'), a position where lysine is predominantly deprotonated ($pK_a \approx 5.3$; Fig. 2) and where proton binding–unbinding events take place most slowly (Table 1). Although both arginine and lysine at this position block the current to similar extents (extent of channel block, ~ 0.3), the 11' arginine remained charged. Excursions to a deprotonated state could not be detected even at pH 9.0. We thus suggest that, at physiological pH, engineered arginines would keep the proton bound at all positions of the stretch between -4' and 17' with the likely exception of positions 0' and 4' (see above), where the arginine would be completely deprotonated despite its high proton affinity in bulk water.

Figure 2 and Table 1 (see also Supplementary Fig. 2) summarize the pK_a results for the δ -subunit. Similar analysis on the α -, β - and ϵ -subunits suggests that, albeit with some clear differences, this general pattern of ΔpK_a values is conserved in all four subunits (data not shown).

pK_a shifts and transfer free-energy changes

The different proton affinities of an ionizable residue in a protein and in bulk water are related, through thermodynamic cycles, to the changes in free energy associated with the transfer of the protonated versus deprotonated forms of the residue from bulk water to its position in the protein ($\Delta\Delta G_{\text{bulk} \rightarrow \text{protein}}^0 \approx -1.36\Delta pK_a \text{ kcal mol}^{-1}$ at 22 °C). Hence, the ΔpK_a value is a measure of the extent to which interactions with the protein microenvironment (water molecules, protein dipoles, charges on ionizable residues) compensate for the loss of solvation free energy incurred on removing the charge of the residue from bulk water.

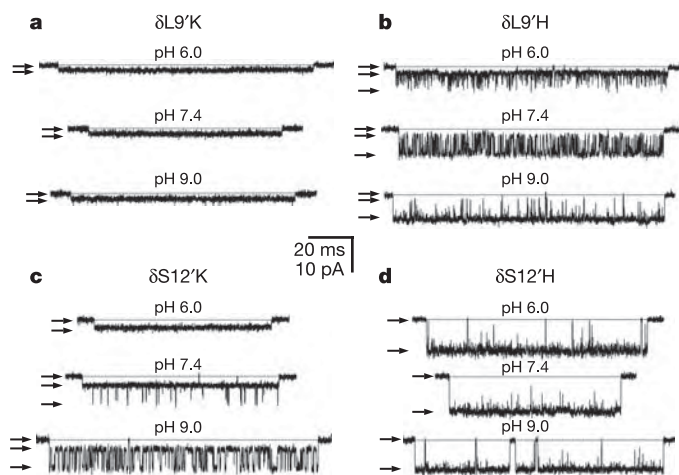


Figure 3 | pH dependence of proton-transfer reactions. **a, b,** Examples of single-channel inward currents ($V \approx -100$ mV, $1 \mu M$ ACh) recorded from leucine to lysine (**a**) and leucine to histidine (**b**) 9' mutants of the δ -subunit ($\delta 9'$) at three different pH values. Because the microenvironment around $\delta 9'$ slightly stabilizes the positive ϵNH_3^+ charge relative to bulk, a lysine remains protonated in the pH range 6.0–9.0. Thus, a histidine, with a lower proton affinity, is needed to reveal the ΔpK_a of this position. As expected, the deprotonated state of the histidine is favoured as the pH increases. **c, d,** Examples of single-channel inward currents ($V \approx -100$ mV; ACh, $1 \mu M$) recorded from serine to lysine (**c**) and serine to histidine (**d**) 12' mutants at three different pH values. Because the microenvironment around $\delta 12'$ destabilizes the positive charge relative to bulk, a histidine remains deprotonated in the pH range 6.0–9.0. By contrast, an arginine at this position retains the labile proton, even at pH 9.0 (data not shown). Thus, only a lysine, with an intermediate proton affinity, can reveal the ΔpK_a at $\delta 12'$. As expected, the deprotonated state of the lysine is favoured as the pH increases. Openings are downwards. Arrows indicate the shut and the two open current levels. The shut level is also indicated by a horizontal broken line. Display $f_c = 6$ kHz.

Table 1 | Position dependence of rates of proton transfer and pK_a values at -100 mV

Position probed	Mutation	pH*	Deprotonation rate (s^{-1})	Protonation rate (s^{-1})	$pK_{a,pore}$	$\Delta pK_a^\dagger$	Intervals ‡	No. $\S$
$\delta - 4'$	C $\rightarrow$ H	7.4	$5,892 \pm 1,162$	$2,859 \pm 549$	7.09 ± 0.003	+0.69	119,772	2
$\delta - 2'$	G $\rightarrow$ H	7.4	$2,913 \pm 140$	$26,540 \pm 2,339$	8.36 ± 0.06	+1.96	111,715	2
$\delta - 1'$	E $\rightarrow$ H	7.4	$16,257 \pm 15$	$1,063 \pm 17$	6.22 ± 0.008	-0.18	52,660	2
$\delta 0'$	K $\rightarrow$ A,R	6.0	—	—	<5.0¶	<-7.00	—	—
$\delta 1'$	T $\rightarrow$ K	7.4	413 ± 32	$4,252 \pm 462$	8.41 ± 0.08	-1.99	74,559	6
$\delta 2'$	S $\rightarrow$ H	7.4	$5,537 \pm 87$	$3,126 \pm 56$	7.15 ± 0.01	+0.75	111,071	2
$\delta 3'$	V $\rightarrow$ K	6.0	$18,045 \pm 937$	$2,434 \pm 138$	5.13 ± 0.005	-5.27	32,384	4
$\delta 4'$	A $\rightarrow$ R	6.0	—	—	<5.0¶	<-7.00	—	—
$\delta 5'$	I $\rightarrow$ H	7.4	$15,829 \pm 620$	$3,331 \pm 105$	6.72 ± 0.003	+0.32	48,977	2
$\delta 6'$	S $\rightarrow$ H	7.4	$7,499 \pm 297$	$5,970 \pm 104$	7.30 ± 0.01	+0.90	549,337	3
$\delta 7'$	V $\rightarrow$ K	6.0	$12,726 \pm 18$	$4,068 \pm 44$	5.50 ± 0.005	-4.90	15,155	2
$\delta 8'$	L $\rightarrow$ K	7.4	740 ± 80	160 ± 9	6.74 ± 0.05	-3.66	66,552	3
$\delta 9'$	L $\rightarrow$ H	7.4	$7,217 \pm 141$	$3,646 \pm 108$	7.10 ± 0.004	+0.70	465,715	2
$\delta 10'$	A $\rightarrow$ K	7.4	$1,792 \pm 77$	$2,383 \pm 63$	7.53 ± 0.01	-2.87	155,089	8
$\delta 11'$	Q $\rightarrow$ K	6.0	85 ± 7	16 ± 1	5.27 ± 0.02	-5.13	12,473	2
$\delta 12'$	S $\rightarrow$ K	7.4	443 ± 11	$13,028 \pm 345$	8.87 ± 0.008	-1.53	69,745	4
$\delta 13'$	V $\rightarrow$ H	7.4	$19,102 \pm 422$	$3,655 \pm 294$	6.68 ± 0.03	+0.28	541,227	3
$\delta 14'$	F $\rightarrow$ K	7.4	$1,799 \pm 90$	395 ± 12	6.74 ± 0.03	-3.66	41,793	4
$\delta 15'$	L $\rightarrow$ H	6.0	$8,173 \pm 518$	$3,835 \pm 314$	5.67 ± 0.008	-0.73	15,245	2
$\delta 16'$	L $\rightarrow$ H	6.0	—	—	6.7☆	+0.30	—	—
$\delta 17'$	L $\rightarrow$ H	6.0	—	—	6.4☆	0	—	—

*The pH value used in calculating the pK_a . At several positions, proton-transfer events were clearly resolved only at pH 6.0 (in the pipette). Although pK_a values turned out to be slightly pH dependent (data not shown), the deviations from their values at pH 7.4 are expected to be small.

$^\dagger \Delta pK_a = pK_{a,pore} - pK_{a,bulk}$.

‡ Intervals indicates the total number of idealized intervals used for kinetic analysis.

$\S$ No. indicates the number of independent patches. Standard errors and means were calculated from the means of individual patches.

||AChRs bearing these δ -subunit mutations also contained the T264P mutation in the non-adjacent ϵ -subunit (M2 $\epsilon 12'$ position). This mutation prolongs individual activations of the channel⁴² and thus increases the number of main-level to sublevel transitions recorded.

¶ pK_a values estimated qualitatively on the basis of the finding that arginines engineered at these positions seem to be completely deprotonated, even at pH 6.0.

☆Although the protonation-deprotonation of histidines engineered at these positions was evident from the electrophysiological recordings, the individual transitions could not be resolved reliably. Thus, these pK_a values are qualitative estimates made on the basis of a comparison with the appearance of recordings from the other constructs.

Because naturally occurring ionizable residues are present only at the ends of M2, and are absent from the membrane-spanning portions of M1, M3 and M4 (Supplementary Fig. 1a, b), the contribution of charge-charge interactions to this energetic balance is expected to be small for most positions examined here. Instead, most of the ΔpK_a values reported in Fig. 2 and Table 1 are expected to be largely governed by the interactions of the positive charges on the engineered basic residues with protein dipoles (backbone dipoles, polar side chains) and water molecules in the heterogeneous environment surrounding M2. An inspection of the colour-coded ΔpK_a maps in Fig. 2 suggests that the observed 'gradient' of pK_a shifts specifically reflects the unique distribution of water molecules around the M2 α -helix of the δ -subunit.

The gating conformational change

Despite significant steps towards elucidating the atomic resolution structure of the full receptor²³, and towards an understanding of some general aspects of the chemical dynamics of the gating reaction^{28–30}, several basic aspects of the AChR remain unresolved, such as the exact sides of the M2 α -helices that face the lumen of the pore in the closed and open states^{23,24,31,32}, the structural rearrangements underlying the closed to open transition^{23,24,33}, and the locations and modes of operation of the activation^{23,34} and desensitization gates. Our experimental data unambiguously define the side of the δ -subunit M2 α -helix that faces the aqueous lumen of the pore in the open state (Figs 1d and 2b, c). Because our approach probes the effect of protonation-deprotonation on single-channel conductance, the proton-transfer rates in the closed channel (and, by inference, the corresponding orientation of the α -helices) were not investigated here.

Other approaches, however, have been used to identify residues that line the pore of the closed channel. Photoincorporation of affinity labels, for example, has provided compelling evidence that positions 9', 13' and 16' face the closed-channel lumen in a muscle-type AChR^{35,36}. A comparison with Figs 1d and 2b, c thus reveals that the inner lining of the pore of the AChR is nearly the same in the closed and open states. Together, these data are consistent with a mechanism of channel opening that simply involves a widening of

the pore, in which the narrowest constriction switches from a location near the middle of the membrane²³ to a location near the intracellular entrance (positions $-2'$ and $-1'$; Fig. 1d). The rotation of M2, if any, is minimal (see also Supplementary Discussion).

Naturally occurring ionizable residues

The protonation state of wild-type residues is another vexing issue. Indeed, the net charge of the rings of ionizable residues that flank the pore of Cys-loop receptors³⁷ and voltage-dependent Na^+ and Ca^{2+} channels³⁸, for example, and the specific charge-stabilizing interactions that keep the pK_a of the voltage-sensing S4 arginines well above physiological pH, remain elusive. On the basis of the insight gained here, the pK_a values of some of the native ionizable groups of the AChR, and other members of the superfamily, can be predicted (see Supplementary Discussion).

Concluding remarks

Gratifyingly, both the channel-block data (Fig. 1d) and the proton-affinity measurements (Fig. 2b, c) are highly consistent with each other and with the proposed α -helical secondary structure of M2 in the closed state^{23,35,39}. Other labelling methods have also been used to explore the properties of the open-channel pore of the AChR but, because in the presence of agonist the channel interconverts rapidly among many different allosteric (closed, open and various desensitized) states, labelling results can seldom be ascribed totally to the open-channel conformation. Because identification of the open state in a single-channel recording is unambiguous (Figs 1b and 3), our approach circumvents this crucial problem altogether.

Here, application of this method has allowed us to infer structural information on the open-channel pore of the AChR with unprecedented detail and precision, to suggest protonation states for some of the naturally occurring ionizable residues, to attain a deeper understanding of the dielectric properties of the pore, and to provide an extensive set of highly shifted pK_a values, which, as the resolution of structural models of the AChR improves, could be used as meaningful benchmarks to validate theoretical electrostatic models for ion-channel proteins. More broadly, these data remind us that basic and acidic amino acids are not charged but chargeable residues, the

pK_a values of which are complex functions of the local microenvironment, and that the high energetic cost of either burying the charged form of a residue or exposing it to lipids can be relieved by simply releasing (basic residues) or taking (acidic ones) a proton. We anticipate that many more facets of the relationship among structure, function and electrostatics in ion channels will be illuminated by the application of this experimental approach.

METHODS

General procedures. HEK-293 cells were transiently transfected with mouse-muscle adult-type AChR complementary DNAs. Mutations were engineered with a QuikChange site-directed mutagenesis kit (Stratagene) and were confirmed by dideoxy sequencing. Single-channel patch-clamp currents were recorded in the cell-attached configuration at 22°C. The bath solution was (in mM): 142 KCl, 5.4 NaCl, 1.8 CaCl₂, 1.7 MgCl₂ and 10 HEPES/KOH buffer (pH 7.4). The pipette solution was identical to that in the bath with the exception of the H⁺ buffer, which changed depending on the desired pH of the solution. These buffers were acetic acid/acetate (pH 5.0), MES (pH 6.0), HEPES (pH 7.4), TABS (pH 9.0) and CAPS (pH 10.5), all titrated to a final pH with KOH. The pipette solution also contained 1 μ M acetylcholine (ACh).

Kinetic analysis. Protonation and deprotonation rates, as well as all other transition rates, were estimated from maximum-likelihood fitting of dwell-time series to kinetic models based on that in Fig. 2a (see Supplementary Methods for more details). To this end, we used QuB software^{40,41} with a retrospectively imposed time resolution of 25 μ s. At the low concentration of ACh used in the experiments described here (1 μ M), only the protonation-deprotonation rates (open $\rightarrow$ open-H⁺ and open-H⁺ $\rightarrow$ open), and the channel-shutting rates (open-H⁺ $\rightarrow$ shut-H⁺ and open $\rightarrow$ shut) can be ascribed a clear physical meaning.

Received 13 June; accepted 29 September 2005.

- Perutz, M. F. Electrostatic effects in proteins. *Science* **201**, 1187–1191 (1978).
- Warshel, A. Electrostatic basis of structure–function correlation in proteins. *Acc. Chem. Res.* **14**, 284–290 (1981).
- Davis, M. E. & McCammon, J. A. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* **90**, 509–521 (1990).
- Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).
- Imoto, K. *et al.* Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. *Nature* **335**, 645–648 (1988).
- Kelley, S. P., Dunlop, J. I., Kirkness, E. F. & Lambert, J. J. A cytoplasmic region determines single-channel conductance in 5-HT₃ receptors. *Nature* **424**, 321–324 (2003).
- Heinemann, S. H., Terlau, H., Stühmer, W., Imoto, K. & Numa, S. Calcium channel characteristics conferred on the sodium channel by single mutations. *Nature* **356**, 441–443 (1992).
- Yang, J., Ellinor, P. T., Sather, W. A., Zhang, J. F. & Tsien, R. W. Molecular determinants of Ca²⁺ selectivity and ion permeation in L-type Ca²⁺ channels. *Nature* **366**, 158–161 (1993).
- Keramidas, A., Moorhouse, A. J., Pierce, K. D., Schofield, P. R. & Barry, P. H. Cation-selective mutations in the M2 domain of the inhibitory glycine receptor channel reveal determinants of ion-charge selectivity. *J. Gen. Physiol.* **119**, 393–410 (2002).
- Lu, Z. & MacKinnon, R. Electrostatic tuning of Mg²⁺ affinity in an inward-rectifier K⁺ channel. *Nature* **371**, 243–246 (1994).
- Khan, A., Romantseva, L., Lam, A., Lipkind, G. & Fozzard, H. A. Role of outer ring carboxylates of the rat skeletal muscle sodium channel pore in proton block. *J. Physiol. Lond.* **543**, 71–84 (2002).
- Schulte, U. & Fakler, B. Gating of inward-rectifier K⁺ channels by intracellular pH. *Eur. J. Biochem.* **267**, 5837–5841 (2000).
- Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K⁺ channel. *Neuron* **16**, 1169–1177 (1996).
- Seoh, S.-A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K⁺ channel. *Neuron* **16**, 1159–1167 (1996).
- Schutz, C. N. & Warshel, A. What are the dielectric ‘constants’ of proteins and how to validate electrostatic models? *Proteins* **44**, 400–417 (2001).
- Mehler, E. L., Fuxreiter, M., Simon, I. & Garcia-Moreno, E. B. The role of hydrophobic microenvironments in modulating pK_a shifts in proteins. *Proteins* **48**, 283–292 (2002).
- Parzefall, F., Wilhelm, R., Heckmann, M. & Dudel, J. Single channel currents at six microsecond resolution elicited by acetylcholine in mouse myoballs. *J. Physiol. Lond.* **512**, 181–188 (1998).
- Dao-pin, S., Anderson, D. E., Baase, W. A., Dahlquist, F. W. & Matthews, B. W. Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry* **30**, 11521–11529 (1991).
- Prod'hom, B., Pietrobon, D. & Hess, P. Direct measurement of proton transfer rates to a group controlling the dihydropyridine-sensitive Ca²⁺ channel. *Nature* **329**, 243–246 (1987).
- Root, M. J. & MacKinnon, R. Two identical noninteracting sites in an ion channel revealed by proton transfer. *Science* **265**, 1852–1856 (1994).
- Lester, H. The permeation pathway of neurotransmitter-gated ion channels. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 267–292 (1992).
- Unwin, N. Acetylcholine receptor channel imaged in the open state. *Nature* **373**, 37–43 (1995).
- Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955 (2003).
- Horenstein, J., Wagner, D. A., Czajkowski, C. & Akabas, M. H. Protein mobility and GABA-induced conformational changes in GABA_A receptor pore-lining M2 segment. *Nature Neurosci.* **4**, 477–485 (2001).
- Corringer, P. J. *et al.* Mutational analysis of the charge selectivity filter of the $\alpha 7$ nicotinic acetylcholine receptor. *Neuron* **22**, 831–843 (1999).
- Adcock, C., Smith, G. R. & Sansom, M. S. P. Electrostatics and the ion selectivity of ligand-gated channels. *Biophys. J.* **75**, 1211–1222 (1998).
- Stopar, D., Spruijt, R. B., Wolfs, C. J. & Hemminga, M. A. Local dynamics of the M13 major coat protein in different membrane-mimicking systems. *Biochemistry* **35**, 15467–15473 (1996).
- Grosman, C., Zhou, M. & Auerbach, A. Mapping the conformational wave of acetylcholine receptor channel gating. *Nature* **403**, 773–776 (2000).
- Cymes, G. D., Grosman, C. & Auerbach, A. Structure of the transition state of gating in the acetylcholine receptor channel pore: a ϕ -value analysis. *Biochemistry* **41**, 5548–5555 (2002).
- Grosman, C. Free-energy landscapes of ion-channel gating are malleable: changes in the number of bound ligands are accompanied by changes in the location of the transition state in acetylcholine-receptor channels. *Biochemistry* **42**, 14977–14987 (2003).
- Goren, E. N., Reeves, D. C. & Akabas, M. H. Loose protein packing around the extracellular half of the GABA_A receptor $\beta 1$ subunit M2 channel-lining segment. *J. Biol. Chem.* **279**, 11198–11205 (2004).
- Kim, S., Chamberlain, A. K. & Bowie, J. U. A model of the closed form of the nicotinic acetylcholine receptor M2 channel pore. *Biophys. J.* **87**, 792–799 (2004).
- Dahan, D. *et al.* A fluorophore attached to nicotinic acetylcholine receptor $\beta M2$ detects productive binding of agonist to the $\alpha \delta$ site. *Proc. Natl Acad. Sci. USA* **101**, 10195–10200 (2004).
- Wilson, G. G. & Karlin, A. The location of the gate in the acetylcholine receptor channel. *Neuron* **20**, 1269–1281 (1998).
- White, B. H. & Cohen, J. B. Agonist-induced changes in the structure of the acetylcholine receptor M2 regions revealed by photoincorporation of an uncharged nicotinic noncompetitive antagonist. *J. Biol. Chem.* **267**, 15770–15783 (1992).
- Arévalo, E., Chiara, D. C., Forman, S. A., Cohen, J. B. & Miller, K. W. Gating-enhanced accessibility of hydrophobic sites within the transmembrane region of the nicotinic acetylcholine receptor's δ subunit. A time-resolved photolabeling study. *J. Biol. Chem.* **280**, 13631–13640 (2005).
- O'Mara, M., Barry, P. H. & Chung, S. H. A model of the glycine receptor deduced from Brownian dynamics. *Proc. Natl Acad. Sci. USA* **100**, 4310–4315 (2003).
- Corry, B., Vora, T. & Chung, S. H. Electrostatic basis of valence selectivity in cationic channels. *Biochim. Biophys. Acta* **1711**, 72–86 (2005).
- Akabas, M. H., Kaufmann, C., Archdeacon, P. & Karlin, A. Identification of acetylcholine receptor channel-lining residues in the entire M2 segment of the α subunit. *Neuron* **13**, 919–927 (1994).
- Qin, F. Restoration of single-channel currents using the segmental k -means method based on hidden Markov modeling. *Biophys. J.* **86**, 1488–1501 (2004).
- Qin, F., Auerbach, A. & Sachs, F. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* **70**, 264–280 (1996).
- Ohno, K. *et al.* Congenital myasthenic syndrome caused by prolonged acetylcholine receptor channel openings due to a mutation in the M2 domain of the epsilon subunit. *Proc. Natl Acad. Sci. USA* **92**, 758–762 (1995).
- Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Sine for wild-type muscle AChR subunit cDNA; J. Jasielec and J. Gasser for technical assistance; S. Varma and B. Garcia-Moreno E. for discussions; and E. Tajkhorshid for introducing us to the VMD imaging program. This work was supported by a grant from the NIH (to C.G.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.G. (grosman@life.uiuc.edu).

Hypomethylation-linked activation of *PAX2* mediates tamoxifen-stimulated endometrial carcinogenesis

Huijian Wu¹, Yupeng Chen¹, Jing Liang¹, Bin Shi¹, Ge Wu¹, Ying Zhang¹, Dan Wang¹, Ruifang Li¹, Xia Yi¹, Hua Zhang¹, Luyang Sun¹ & Yongfeng Shang¹

Tamoxifen, a selective oestrogen receptor modulator, has been used in the treatment of all stages of hormone-responsive breast cancer. However, tamoxifen shows partial oestrogenic activity in the uterus and its use has been associated with an increased incidence of endometrial cancer. The molecular explanation for these observations is not known. Here we show that tamoxifen and oestrogen have distinct but overlapping target gene profiles. Among the overlapping target genes, we identify a paired-box gene, *PAX2*, that is crucially involved in cell proliferation and carcinogenesis in the endometrium. Our experiments show that *PAX2* is activated by oestrogen and tamoxifen in endometrial carcinomas but not in normal endometrium, and that this activation is associated with cancer-linked hypomethylation of the *PAX2* promoter.

The triphenylethylene derivative tamoxifen has been used as the treatment of choice for all stages of hormone-responsive breast cancer and it can prevent breast cancer in high-risk women¹. However, tamoxifen shows partial oestrogenic effects in other target tissues^{1,2}. These partial oestrogenic actions produce beneficial effects on bones and the cardiovascular system in postmenopausal women but are also associated with an increased incidence of endometrial cancer^{1,2}. Endometrial carcinoma is the most common gynaecological malignancy in which oestrogen has been identified as a classic aetiological factor^{3,4}; clinically, most endometrial cancers are type I oestrogen-dependent endometrioid adenocarcinomas⁵.

Despite well-established epidemiological evidence that oestrogen is involved in endometrial carcinogenesis and the well-documented association of tamoxifen with increased incidence of endometrial cancer, the molecular basis underlying this association is not understood. Oestrogen exerts its biological activities by binding to oestrogen receptors (ERs), ER α and ER β , which function as ligand-dependent transcription factors and regulate target gene transcription. It is thought, on the basis of crystal structures of the hormone-binding domains of ERs bound to ligand^{6,7}, that tamoxifen acts as an ER antagonist by binding to ERs and by inducing a conformational change that blocks the interaction of ERs with coactivator proteins in the mammary gland. This molecular mechanism is not compatible, however, with the partial oestrogenic activity of tamoxifen in the uterus. Relevant to this, we have shown that tamoxifen-bound ER can interact with a p160 family of coactivators⁸, suggesting that tamoxifen may be actively involved in gene regulation in the endometrium.

Genomic action of tamoxifen in endometrium

To understand the genomic basis for the partial agonistic activity of tamoxifen in the endometrium, we first investigated, using Affymetrix human genome arrays, the gene expression profile of endometrial epithelial cells (EECs) immunomagnetically purified from type I endometrioid carcinoma samples (cancerous EECs or cEECs) under

treatment with oestrogen and tamoxifen. cEECs from stage I and stage II endometrioid carcinoma samples were pooled separately and cultured. Pilot experiments were done first to determine the optimal duration of treatment, and we found that a 3-h treatment of cEECs with 17 β -oestradiol (E2) yielded the greatest coverage (about 75%) of currently known ER target genes^{9–12}. Thus, we chose a 3-h treatment for subsequent experiments.

We grew cEECs in the absence of oestrogen and then treated them with E2 or tamoxifen for 3 h or left them untreated (control). Each treatment was done in duplicate, and each replicate was analysed in duplicate arrays. Gene array analysis indicated that oestrogen and tamoxifen regulated the transcription of two sets of genes that had an overlapping but distinct pattern (Fig. 1a). There were 97 genes that were regulated by oestrogen and 114 genes that were regulated by tamoxifen. Among these genes, only 35 were regulated (21 upregulated and 14 downregulated) by both oestrogen and tamoxifen (Supplementary Fig. 2); thus, most of the genes were uniquely regulated by either oestrogen or tamoxifen.

Similar experiments were done in EECs that were immunomagnetically purified from samples of normal endometrium that were age-matched to the cancers (normal EECs or nEECs). The comparison of cEECs with nEECs showed an overall increase in gene expression in cEECs: more genes were upregulated in cEECs by oestrogen (42 versus 33) and tamoxifen (51 versus 35), and fewer genes were downregulated in cEECs by oestrogen (55 versus 62) and tamoxifen (63 versus 70). These experiments clearly show that tamoxifen does not function merely by affecting the transcription of oestrogen target genes. Instead, tamoxifen has its own target genes and these include some of those targeted by oestrogen. Therefore, the genomic view of tamoxifen action should include some of the oestrogen target genes as well as the target genes that are unique to tamoxifen (Supplementary Fig. 2).

We verified the results of the gene array experiments by using real-time polymerase chain reaction with reverse transcription (RT-PCR) to analyse the expression of selected genes in both EEC cultures and

¹Department of Biochemistry and Molecular Biology, Peking University Health Science Center, Beijing 100083, China.

in rat uterine tissues under treatment with oestrogen or tamoxifen. Six genes, *erythroid Kruppel-like factor* (*EKLF*), *protein kinase C α* (*PKC α*), *paired-box gene 2* (*PAX2*), *human kidney water channel* (*hKID*), *cyclin F* and *retinoic acid receptor β 1* (*RAR β 1*), were selected for their functional relevance to cell growth and proliferation and to represent different levels of regulation. Of these genes, *EKLF*, *PKC α* and *PAX2* represented genes that were upregulated by both oestrogen and tamoxifen in EEC cells at high, modest and low levels, respectively; *hKID*, *cyclin F* and *RAR β 1* represented genes that were down-regulated in these cells at high, modest and low levels, respectively. The expression of *EKLF*, *PKC α* and *PAX2* was increased in cEECs by treatment with tamoxifen or oestrogen, whereas that of *hKID*, *cyclin F* and *RAR β 1* was decreased with this treatment (Fig. 1b). The expression pattern of these genes measured by real-time RT-PCR was analogous to that obtained in the gene array analysis.

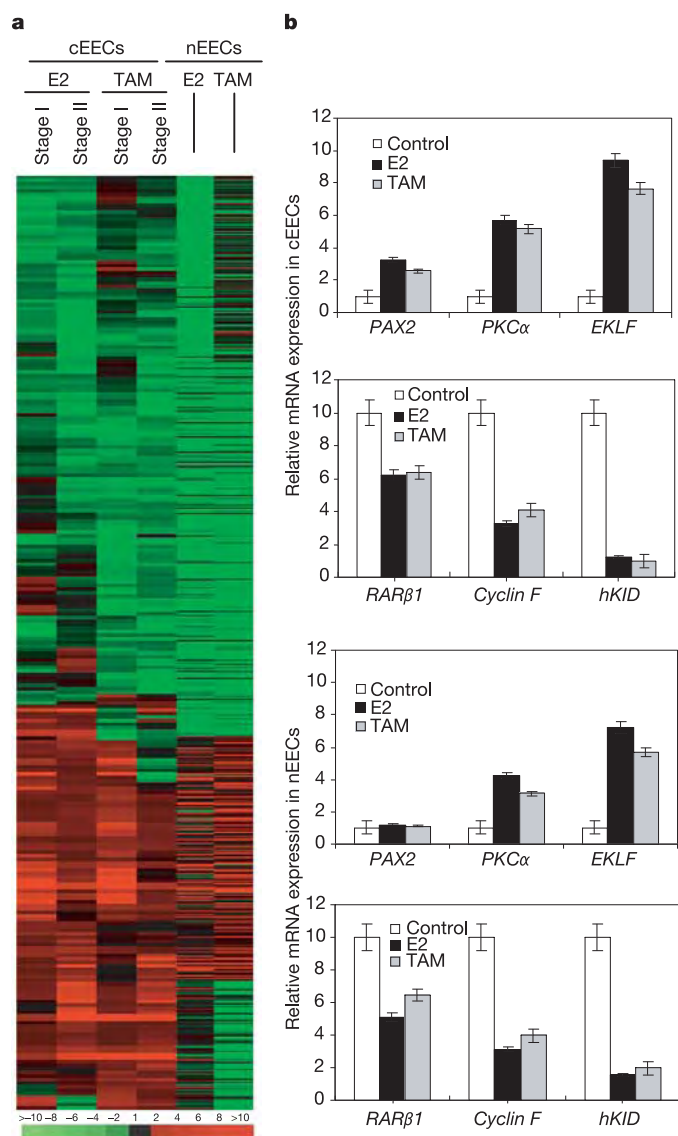


Figure 1 | Genomic view of tamoxifen action. **a**, Clustering of genes regulated by E2 and tamoxifen (TAM) in EECs immunomagnetically purified from stage I and stage II endometrioid carcinoma samples (cEECs) and from samples of normal endometrial epithelium (nEECs). **b**, Real-time RT-PCR verification of the gene array analysis. cEECs and nEECs were untreated (control) or treated with either 100 nM E2 or 5 μ M tamoxifen for 3 h. Trizol Reagent was used to extract total RNAs for analysing mRNA expression by real-time RT-PCR. Each bar represents the mean $\pm$ s.d. of triplicate experiments.

Similar experiments were done in nEECs. In these cells, the expression patterns of *EKLF*, *PKC α* , *hKID*, *cyclin F* and *RAR β 1* were consistent with those observed in cEECs by gene array analysis and real-time RT-PCR analysis, except for *PAX2*, whose expression was not affected in nEECs (Fig. 1b). Similar results were also obtained in rat uterine tissues from ovariectomized adult female CD rats. The proliferative effect of oestrogen and tamoxifen on the endometrium was also confirmed by measuring the wet weight of the treated rat uteri and the thickness of the luminal epithelia, and by immunohistochemical staining of the cell proliferation marker, proliferating cell nuclear antigen, and the epithelial cell marker, Ki-67 (Supplementary Fig. 3).

Molecular effectors of tamoxifen in cancer

Because oestrogen and tamoxifen are both implicated in endometrial carcinogenesis and are both thought to exert their biological activities through binding to ERs, we considered that molecular effectors that mediate the carcinogenic roles of oestrogen and tamoxifen might be among the genes that are commonly regulated by these two molecules. Both gain-of-function and loss-of-function experiments were done on these genes to examine their roles in the proliferation of endometrial carcinoma cells and in the growth stimulation of transplanted endometrial tumours in athymic mice.

First, we reasoned that if the upregulated genes are the effectors of oestrogen and tamoxifen in endometrial cell proliferation, then overexpression of these genes in endometrial carcinoma cells would lead to cell proliferation without stimulation by oestrogen or tamoxifen. Oestrogen-responsive endometrial carcinoma cell lines, ECC-1 (refs 8, 13–15) and Ishikawa^{8,16–18}, were used in these experiments. These cells were infected individually with retroviruses carrying complementary DNAs for the 21 upregulated genes, and the effect of this overexpression on cell proliferation was analysed by flow cytometry.

Both oestrogen and tamoxifen stimulated the proliferation of ECC-1 and Ishikawa cells infected with retroviruses carrying an empty vector (Fig. 2a). Notably, even without stimulation by oestrogen or tamoxifen, ECC-1 and Ishikawa cells infected with retroviruses carrying *PAX2* also showed significant stimulation of cell proliferation. Cells infected with retroviruses carrying *EKLF* or *PKC α* showed marginal stimulation of cell proliferation, whereas cells infected with retroviruses carrying cDNAs for the remainder of the 21 genes showed minimal effects (data not shown). Overexpression of the genes was confirmed by western blotting when antibodies were available or RT-PCR when antibodies were not available (see Fig. 2b for examples of western blotting for *EKLF*, *PKC α* or *PAX2*). Although the overexpression of these proteins may not represent physiological levels, these experiments indicate that *PAX2* could be a key effector in mediating cell proliferation in response to oestrogen and tamoxifen treatment in ECC-1 and Ishikawa cells.

To gain support for this observation, we used retrovirus-delivered short interfering (siRNA) to silence individually the expression of the 21 upregulated genes and then investigated the effect of silencing on cell proliferation by flow cytometry analysis. We reasoned that if a gene functions to mediate the cell proliferative effect of oestrogen and tamoxifen, then a knockdown in the expression of that gene will decrease or diminish the cell proliferative effect of oestrogen or tamoxifen. Whereas silencing of the expression of *EKLF* or *PKC α* had marginal effect on the oestrogen- or tamoxifen-stimulated proliferation of ECC-1 and Ishikawa cells, silencing *PAX2* expression led to a marked decrease in proliferation (Fig. 2c). The effect of silencing expression of the other genes on the proliferation of ECC-1 cells was minimal (data not shown). In addition, a *PAX2* mutant in which the activation or repression domain was deleted had a dominant-negative effect on the *PAX2*-mediated proliferation of ECC-1 cells under treatment with oestrogen and tamoxifen, and this effect could be rescued by wild-type *PAX2* (Supplementary Fig. 4). Together,

these findings support a role for PAX2 in mediating the cell proliferative activity of oestrogen and tamoxifen in endometrial cells. The knockdown of gene expression was confirmed by western blotting or RT-PCR, depending on the availability of individual antibodies. Examples of western blotting for EKLF, PAX2 and PKC α are shown in Fig. 2d.

Analogously, gain-of-function and loss-of-function experiments were done with all of the 14 downregulated genes. We reasoned that if downregulation of one of or some of these genes were required for the cell proliferative effect of oestrogen and tamoxifen, then overexpression of these genes would counteract the oestrogen- or tamoxifen-stimulated effect on the proliferation of ECC-1 cells and silencing of the expression of these genes would result in cell proliferation even without stimulation by oestrogen or tamoxifen. However, both gain-of-function and loss-of-function mutation of each of the 14 genes had a minimal effect on oestrogen- or tamoxifen-stimulated cell proliferation (see Supplementary Fig. 5 for examples of experiments with hKID, cyclin F and RAR β 1 in ECC-1 cells). These findings suggest that the key factors mediating the cell proliferative effects of oestrogen and tamoxifen are not among the genes that are downregulated by these molecules.

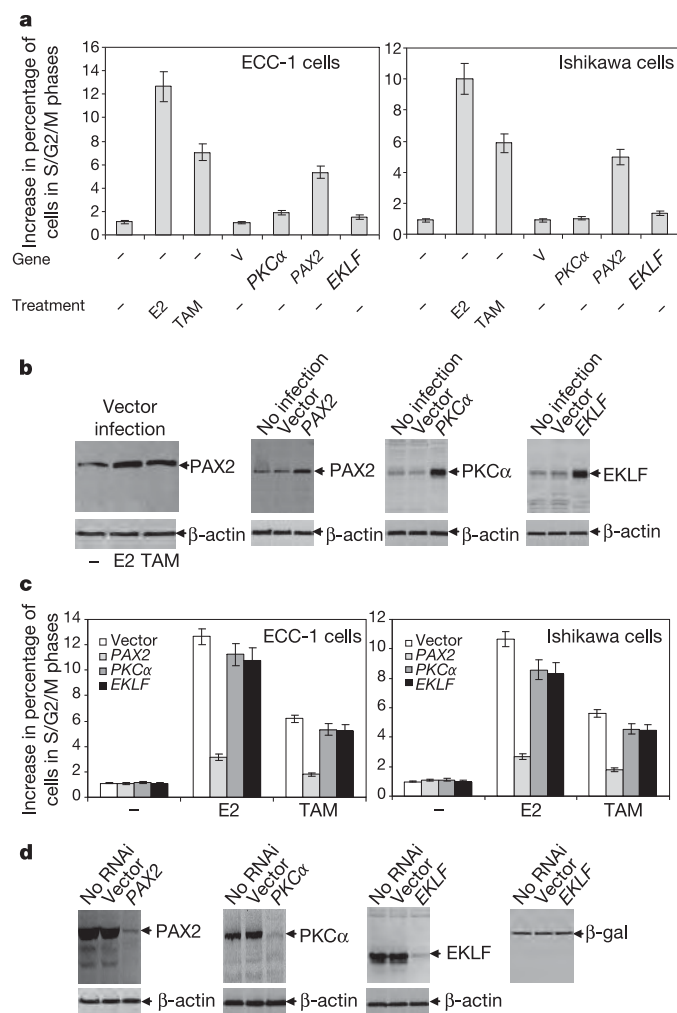


Figure 2 | Effect of EKLF, PKC α and PAX2 on the growth of ECCs.

a, b, ECC-1 and Ishikawa cells were infected with retroviruses carrying EKLF, PKC α or PAX2. Cell proliferation was analysed by FACS (**a**) and protein expression was examined by western blotting (**b**). **c, d**, ECC-1 and Ishikawa cells were infected with retroviruses carrying sequences encoding siRNA specific for EKLF, PKC α or PAX2, and cell proliferation and protein expression were analysed as in **a, b**. In **a** and **c**, each bar represents the mean $\pm$ s.d. of triplicate experiments.

Growth stimulation of ECC-1 tumours by PAX2

To establish further the role of PAX2 in mediating the proliferative effect of oestrogen and tamoxifen and to investigate the possible involvement of PAX2 in endometrial carcinogenesis, we transplanted three types of endometrial tumour developed from ECC-1 cells onto ovariectomized athymic mice (BALB/c; Charles River). The transplanted tumours showed unchanged expression of PAX2 (infected with retroviruses carrying an empty vector), overexpression of PAX2 (infected with retroviruses carrying the PAX2 gene) or specific knockdown of PAX2 expression (infected with retroviruses carrying an oligonucleotide specific for PAX2 siRNA).

Tumour growth was monitored in mice that received no treatment or treatment with oestrogen or tamoxifen. In athymic mice that received a transplant of ECC-1 tumours with unchanged PAX2 expression (neither overexpression nor knockdown), both E2 and tamoxifen stimulated tumour growth over 24 weeks: a greater stimulation was observed with E2 treatment, and a lesser effect was observed with tamoxifen treatment (Fig. 3a). In athymic mice that received a transplant of ECC-1 tumours that had PAX2 overexpression, tumour growth was observed even with no treatment, and treatment with oestrogen or tamoxifen further enhanced the ECC-1 tumour growth (Fig. 3b). In athymic mice that received tumour

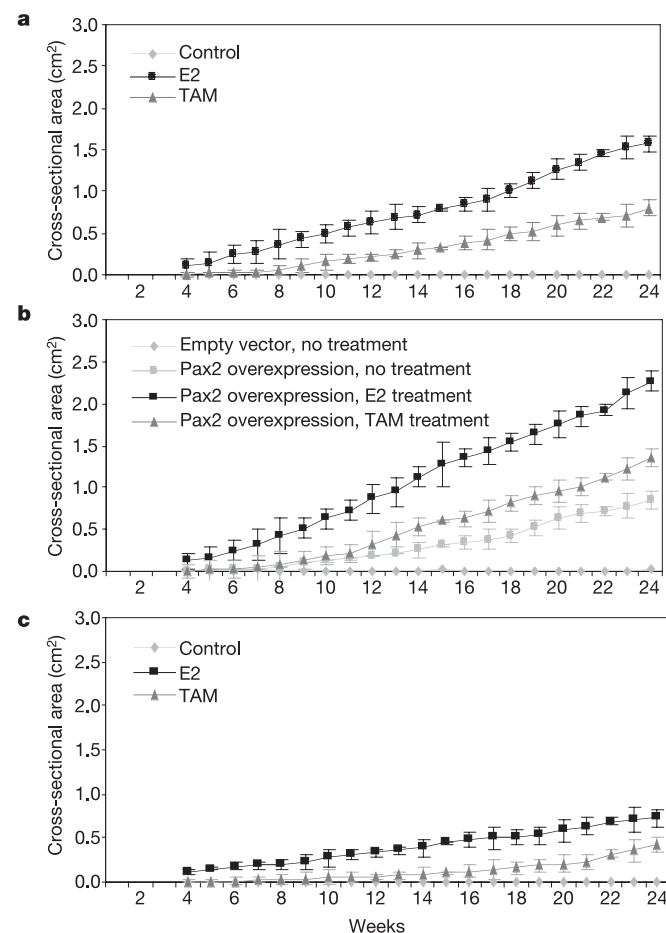


Figure 3 | Effect of PAX2 on the growth of transplanted ECC-1 tumours in nude mice. **a**, No changes in PAX2 expression; **b**, PAX2 overexpression; **c**, PAX2 'knockdown'. ECC-1 cells overexpressing PAX2 or with PAX2-specific knockdown were transplanted onto ovariectomized athymic mice. Tumour growth was monitored after treatment with vehicle, E2 or tamoxifen. ECC-1 cells stably infected with retroviruses carrying an empty vector were used as a control. Tumours were measured weekly with Vernier callipers. The cross-sectional area was calculated using the formula length $\times$ width/4 $\times$ π ; each point represents the mean $\pm$ s.d. of triplicate measurements.

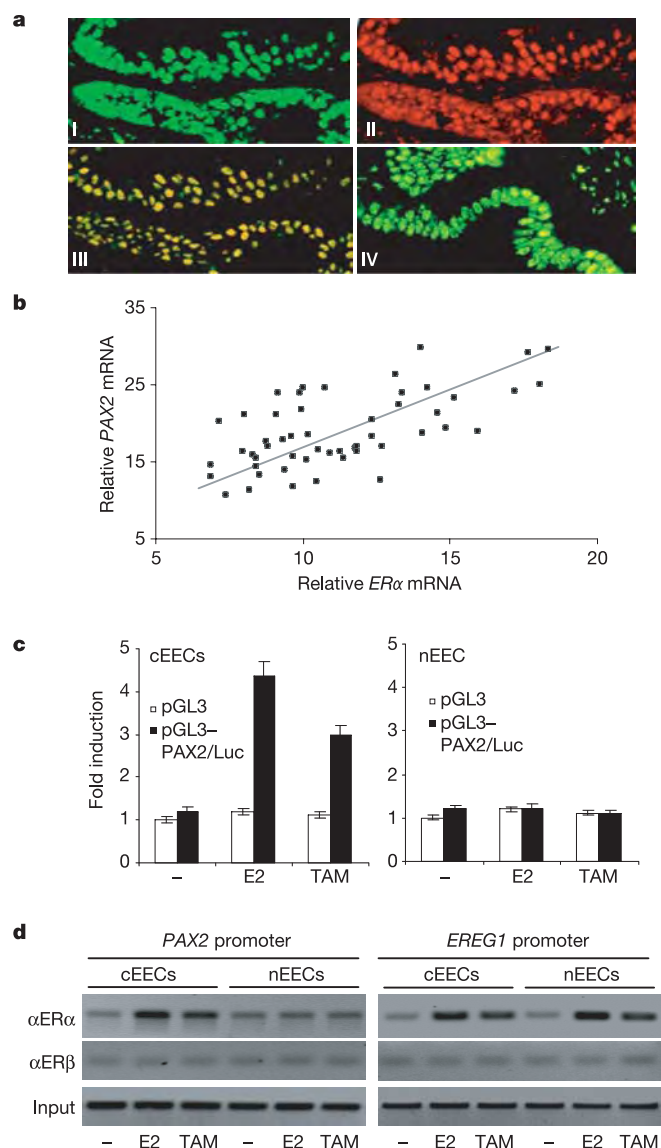


Figure 4 | PAX2 as a downstream target of ER α . **a**, Sections of human endometrial cancer samples were stained for ER α (green, I), PAX2 (red, II) or both (III). Sections adjacent to tumour samples were stained for both ER α and PAX2 (IV). **b**, Real-time RT-PCR analysis of ER α and PAX2 mRNA expression in tumour samples. **c**, Luciferase activity in cEECs (left) and nEECs (right) transfected with a construct encoding luciferase driven by a 2.0-kb upstream regulatory region of PAX2. Each bar represents the mean $\pm$ s.d. of triplicate experiments. **d**, ChIP analysis of the occupancy of ERs on PAX2 and EREG1 promoters.

transplants with PAX2 knockdown, tumour growth stimulation by oestrogen and tamoxifen was greatly attenuated (Fig. 3c).

Similar experiments were done in mice that received transplanted endometrial tumours in which the expression of *EKLF* had been manipulated. Tumour growth in these mice was similar to controls (data not shown). Collectively, these experiments strongly indicate that PAX2 is a key effector of the oestrogen- and tamoxifen-stimulated growth of ECC-1 tumours.

PAX2 as a downstream target of ER α

To support further the observation that PAX2 is a downstream effector of oestrogen and tamoxifen in stimulating the growth of endometrial cells, we analysed the expression of ER α and PAX2 in the 53 ER-positive endometrioid carcinoma samples used in microarray analysis. The expression patterns of ER α and PAX2 proteins were

determined with dual immunofluorescence staining, and the expression levels of ER α and PAX2 messenger RNAs were measured by real-time RT-PCR in these samples. The pattern of ER α expression was well correlated with the expression of PAX2 in endometrial cancer tissue (Fig. 4a). In real-time RT-PCR experiments, the relative level of ER α expression was plotted against that of PAX2 expression (Fig. 4b). Statistical analysis with the SAS system found a Spearman correlation coefficient of 0.71223 ($P < 0.0001$) and a Kendall Taub correlation coefficient of 0.62345 ($P < 0.0001$), further indicating a strong correlation between the expression of ER α and the expression of PAX2 in ER-positive endometrial carcinomas and supporting the idea that PAX2 is a downstream target of ER α in endometrial cells.

Examination of the PAX2 promoter did not reveal a typical oestrogen response element. However, the upstream regulatory region of PAX2 does contain numerous *cis*-elements for binding transcription factors such as Sp1 and NF- κ B¹⁹, which are implicated in indirect binding to ER^{8,20–23}. To gain support for the notion that PAX2 is a direct downstream target of ERs, we investigated the effect of oestrogen and tamoxifen stimulation on the activation of luciferase driven by the PAX2 promoter by using a plasmid construct carrying a ~2.0-kb fragment of the PAX2 promoter plus the first exon¹⁹. Both oestrogen and tamoxifen could indeed stimulate luciferase gene expression, but only in cancerous EECs and not in normal EECs (Fig. 4c). Furthermore, chromatin immunoprecipitation (ChIP) assays detected the presence of ER α but not ER β on the PAX2 upstream regulatory region after treatment with E2 or tamoxifen in cEECs but not in nEECs (Fig. 4d). The ChIP results were confirmed by real-time PCR quantification (Supplementary Fig. 6). Collectively, these findings support the idea that PAX2 is a downstream target of oestrogen and tamoxifen in the endometrium.

PAX2 promoter hypomethylation

To understand why PAX2 is activated in cancerous EECs and cancer cell lines but not in normal endometrium, we examined the methylation status of PAX2 promoter by methylation-specific PCR in the 53 type I endometrioid carcinoma samples and in the 19 samples from normal endometrium. We found that the PAX2 promoter was hypermethylated in normal endometrium, but over 75% of the carcinoma samples (40 of 53) showed PAX2 promoter hypomethylation (Fig. 5a). These results were confirmed by bisulphite DNA sequencing (Fig. 5b). Furthermore, treatment of nEECs with the methyltransferase inhibitor 5-aza-deoxycytidine (5-aza-dC) led to PAX2 reactivation, ER α recruitment and PAX protein expression (Fig. 5c). Together, these results suggest that the reactivation of PAX2 expression in cancerous endometrial cells is associated with loss of the methylation mark in the PAX2 promoter.

Finally, we investigated the possible mechanism underlying the alteration in PAX2 methylation in cancerous versus normal endometrial cells. It is thought that methylated CpG may be bound by methyl-CpG binding proteins (MeCPs), which in turn are associated with transcription repression complexes containing mSin3A and histone deacetylase (HDAC)²⁴. Thus, we examined the association of MeCP2, mSin3A and HDAC1 with the PAX2 upstream regulatory region in nEECs and cEECs by ChIP. We found that MeCP2, mSin3A and HDAC1 were present in the PAX2 upstream regulatory region in nEECs but not in cEECs (Fig. 5d), suggesting that loss of the methylation mark in PAX2 in cEECs is linked to loss of the association of a protein complex containing MeCP2, mSin3A and HDAC1.

Discussion

Biochemical and animal experiments^{6,14,16,17,25–33}, as well as genetic studies^{34–37}, strongly suggest that an ER-dependent pathway implicated in gene transcriptional regulation underlies the mechanism of tamoxifen action in the uterus. We have shown not only that tamoxifen regulates gene transcription in endometrial epithelial

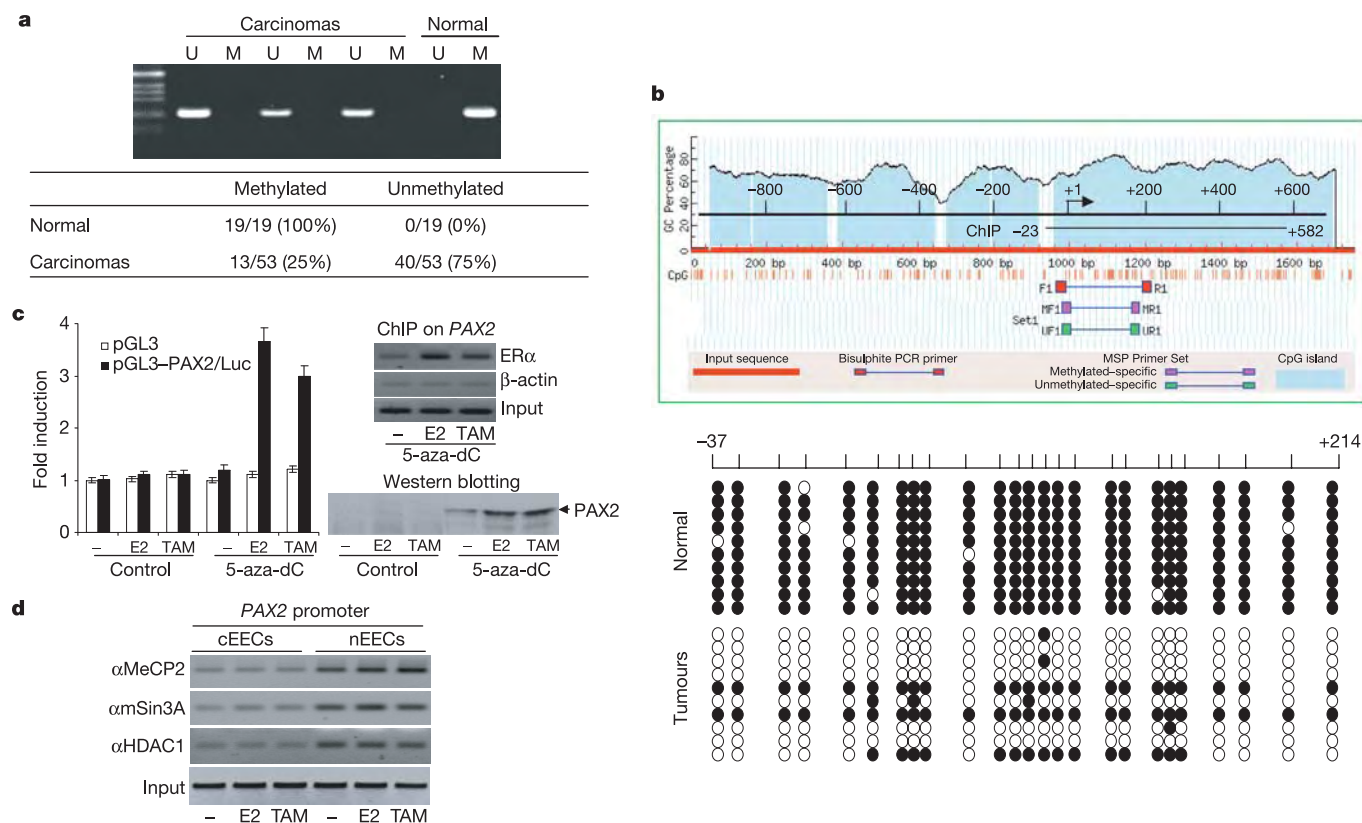


Figure 5 | Cancer-linked hypomethylation of the *PAX2* promoter.

a, Methylation-specific PCR analysis of the *PAX2* upstream regulatory region in endometrial carcinomas and in normal endometrium. M indicates hypermethylated *PAX2*; U indicates unmethylated *PAX2*. **b**, Results of bisulfite DNA sequencing of the *PAX2* upstream regulatory region. **c**, Normal EECs were treated with 5-aza-2'-deoxycytidine and then

transfected with a construct encoding luciferase driven by the upstream regulatory region of *PAX2*. Cells were assayed for luciferase activity, ERα recruitment and *PAX2* expression. Each bar represents the mean $\pm$ s.d. for triplicate experiments. **d**, MeCP2, mSin3A and HDAC1 are associated with the *PAX2* upstream regulatory region in nEECs but not in cEECs, as measured by ChIP assays.

cells, but also that the genes targeted by tamoxifen are largely different from those targeted by oestrogen. Our experiments indicate that gene transcriptional regulation may dictate the role of tamoxifen in endometrial carcinogenesis. Our observations also indicate that tamoxifen is a compound with distinct genomic activity and is not simply a partial ER agonist as traditionally described.

We have shown that *PAX2* is a common target of oestrogen- and tamoxifen-bound ERα and can promote, both *in vitro* and *in vivo*, the growth of endometrial cancer cells. Typically, *PAX2* expression accompanies high rates of cell division^{38–40}. *PAX2* is expressed in Wilms tumour⁴¹, a childhood renal tumour of embryonic origin, and in a high proportion of primary tumours including breast, ovarian, lung, colon, prostate and lymphoma^{39,40,42}. Our observation that *PAX2* is a crucial effector in mediating the proliferation and growth of endometrial cells and tumours also indicates that *PAX2* has a role in carcinogenesis. Furthermore, we found that *PAX2* is silenced in normal endometrium and reactivated in endometrial cancer, and this reactivation is associated with cancer-linked hypomethylation of the *PAX2* promoter. Although hypomethylation was the first epigenetic alteration to be characterized in cancer⁴³, its role has been underappreciated for many years in favour of hypermethylation. However, gene reactivation by cancer-linked hypomethylation has been rediscovered^{44–47}. Future studies need to focus on delineating the cellular milieu surrounding, and the molecular mechanism underlying, both aberrant methylation patterns.

Identification of tumour-specific molecules that function as targets is crucial for the development of cancer drugs and thus is a chief goal of cancer research. Here, the identification of *PAX2* as a target of oestrogen and tamoxifen that mediates their carcinogenic

roles in the uterus may provide useful information for designing safer drugs for the treatment of breast cancer and endometrial cancer. It will be interesting to investigate the mechanism involved in the loss of *PAX2* methylation mark in endometrial carcinomas. Perhaps more relevant to our findings, exactly how *PAX2* functions to promote cell proliferation is still not known, and the pathophysiological relevance and the epigenetic regulation of other oestrogen and tamoxifen target genes need to be explored.

METHODS

Immunomagnetic purification of EECs. Minced endometrial tissues were digested in DMEM/F12 medium (Life Technologies) supplemented with 1% fetal bovine serum, 2 mg ml⁻¹ of collagenase I (Sigma) and 2 mg ml⁻¹ of hyaluronidase (Sigma) at 37°C for 2 h. Cells were collected by centrifugation, digested with trypsin, resuspended in PBS, 1% bovine serum albumin and 2 mM EDTA, and purified by using an Epithelial Enrich kit (Dyna) in accordance with the manufacturer's recommendations. The pooled EEC pellets were resuspended in phenol-red-free DMEM/F12 medium containing 15 mM HEPES, 2.0 nM L-glutamine and 10% charcoal-dextran-stripped calf serum (Hyclone Labs). The purity of the cell cultures was judged to be >96% by immunohistochemical staining and cell sorting using antisera to cytokeratins 7, 10 and 18 (MN 116; Dako).

Human genome array analyses. Total RNA was extracted with Trizol Reagent (Invitrogen) and processed for use on Affymetrix U95A Human Genome arrays according to the manufacturer's protocol. Arrays were scanned with a GeneArray Scanner (Agilent Technologies). Raw data were collected and analysed by Microarray Suite and Data Mining Tools software (Affymetrix). Experiments were done on two replicate chips for each treatment in each ECC cell pool. We used Mann–Whitney pairwise comparisons to identify genes that were differentially expressed. Genes with concordance exceeding 80.6% were considered significantly different ($P < 0.1$). Gene lists from both comparisons were

processed by Genespring software (Silicon Genetics) to identify common Affymetrix probe sets in both lists.

Gene overexpression by a retroviral vector system. We prepared recombinant retroviruses expressing all 35 genes by using a Retro-X Universal Packaging System (BD Biosciences) according to the vendor's manual. mRNA sequences representing the 35 genes and information on their open reading frames (ORFs) were obtained from RefSeq (<http://www.ncbi.nlm.nih.gov/RefSeq/>). The ORFs were individually amplified by PCR and cloned. For infection, cells were incubated with viruses at a multiplicity of infection of 10 for 6 h. Fresh medium was then added and the cultures were continued for another 48 h. To generate PAX2-overexpressing ECC-1 cells for transplantation into nude mice, infected cells were grown in the presence of $400 \mu\text{g ml}^{-1}$ of G418 for 3 weeks. Individual drug-resistant clones were collected, pooled and expanded.

Gene silencing by a retroviral siRNA system. Recombinant retroviruses carrying specific oligonucleotides for generating siRNAs to the 35 genes described here were prepared by using a Knockout RNAi system (BD Biosciences) according to the vendor's manual. mRNA sequences and ORF information were obtained from RefSeq. Sequences were masked to remove repetitive sequences with RepeatMasker (<http://www.repeatmasker.org/>), and vector contamination was masked by searching with NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) against UniVec (<http://www.ncbi.nih.gov/VecScreen/UniVec.html>). Three unique 19-oligonucleotides for each target were selected as described in the Supplementary Methods. For infection, cells were incubated with viruses at a multiplicity of infection of 10 for 6 h. Fresh medium was then added and the cultures were continued for another 48 h. To generate PAX2-silenced ECC-1 cells for transplantation into nude mice, infected cells were grown in the presence of $2 \mu\text{g ml}^{-1}$ of puromycin for 3 weeks. Individual drug-resistant clones were collected, pooled and expanded. The oligonucleotide sequences synthesized for RNA interference are given in the Supplementary Information.

Methylation-specific PCR. We analysed 53 type I endometrioid cancer samples and 19 normal endometrial samples. DNA was extracted and denatured with NaOH and treated with sodium bisulphite for 16 h. The following methylation-specific primers for the PAX2 promoter were used: left M primer, 5'-GGGTTTTTTCGTCGAAGTTC-3'; right M primer, 5'-ACTAAACCTCG ACTCCCGAT-3'; left U primer, 5'-GGTTTTTTTGTGTAAGTTTGG-3'; right U primer, 5'-AAACTAAACCTCAACTCCCAAT-3'. DNA from peripheral blood lymphocytes of healthy individuals and water blanks were used as negative controls for methylated genes. DNA from peripheral blood lymphocytes treated with SssI methyltransferase (New England Biolabs) was used as a positive control for methylated alleles.

Bisulphite sequencing. Bisulphite sequencing was done in 10 tumour samples and 10 normal tissues. Primer sequences for bisulfite sequencing of the PAX2 fragment were 5'-GTTTTGTAGTTTTAGAGAGATATATAT-3' (forward) and 5'-AAATTAACAAAAAATAACAATCCC-3' (reverse), which amplify the -37 to +214 region upstream of PAX2. This area contains 25 CpG sites. Amplified PCR products were purified with a Gel Extraction kit (Qiagen) and ligated into the pCR4-TOPO plasmid vector with a TA-cloning system (Invitrogen). Plasmid-transformed *Escherichia coli* were cultured and plasmid DNA was isolated with QIAprep 96 (Qiagen). Purified plasmid DNA containing the PAX2 sequence was sequenced with an ABI 377 automated sequencer using BigDye Terminator chemistry (Applied Biosystems) and the M13 reverse primer. Samples that had clones with >50% methylation of CpGs were designated partially methylated. Samples containing clones with >75% methylation of CpGs sites were designated methylation positive. All other samples were designated methylation negative.

Received 15 July; accepted 8 September 2005.

1. Fisher, B. *et al.* Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J. Natl Cancer Inst.* **90**, 1371–1388 (1998).
2. Jordan, V. C., Gapstur, S. & Morrow, M. Selective estrogen receptor modulation and reduction in risk of breast cancer, osteoporosis, and coronary heart disease. *J. Natl Cancer Inst.* **93**, 1449–1457 (2001).
3. Persson, I. Estrogens in the causation of breast, endometrial and ovarian cancers—evidence and hypotheses from epidemiological findings. *J. Steroid Biochem. Mol. Biol.* **74**, 357–364 (2000).
4. Akhmedkhanov, A., Zeleniuch-Jacquotte, A. & Toniolo, P. Role of exogenous and endogenous hormones in endometrial cancer: review of the evidence and research perspectives. *Ann. NY Acad. Sci.* **943**, 296–315 (2001).
5. Hernandez, E. Endometrial adenocarcinoma: a primer for the generalist. *Obstet. Gynecol. Clin. North Am.* **28**, 743–757 (2001).
6. Shiau, A. K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).

7. Brzozowski, A. M. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753–758 (1997).
8. Shang, Y. & Brown, M. Molecular determinants for the tissue specificity of SERMs. *Science* **295**, 2465–2468 (2002).
9. Lin, C. Y. *et al.* Discovery of estrogen receptor α target genes and response elements in breast tumour cells. *Genome Biol.* **5**, R66 (2004).
10. O'Lone, R., Frith, M. C., Karlsson, E. K. & Hansen, U. Genomic targets of nuclear estrogen receptors. *Mol. Endocrinol.* **18**, 1859–1875 (2004).
11. Wang, D. Y., Fulthorpe, R., Liss, S. N. & Edwards, E. A. Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. *Mol. Endocrinol.* **18**, 402–411 (2004).
12. Seth, P., Krop, I., Porter, D. & Polyak, K. Novel estrogen and tamoxifen induced genes identified by SAGE (serial analysis of gene expression). *Oncogene* **21**, 836–843 (2002).
13. Ghosh, M. G., Thompson, D. A. & Weigel, R. J. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. *Cancer Res.* **60**, 6367–6375 (2000).
14. Dardes, R. C. *et al.* Regulation of estrogen target genes and growth by selective estrogen-receptor modulators in endometrial cancer cells. *Gynecol. Oncol.* **85**, 498–506 (2002).
15. Farnell, Y. Z. & Ing, N. H. The effects of estradiol and selective estrogen receptor modulators on gene expression and messenger RNA stability in immortalized sheep endometrial stromal cells and human endometrial adenocarcinoma cells. *J. Steroid Biochem. Mol. Biol.* **84**, 453–461 (2003).
16. Sakamoto, T. *et al.* Estrogen receptor-mediated effects of tamoxifen on human endometrial cancer cells. *Mol. Cell. Endocrinol.* **192**, 93–104 (2002).
17. Wang, Z. *et al.* Tamoxifen regulates human telomerase reverse transcriptase (hTERT) gene expression differently in breast and endometrial cancer cells. *Oncogene* **21**, 3517–3524 (2002).
18. Nishida, M. The Ishikawa cells from birth to the present. *Hum. Cell* **15**, 104–117 (2002).
19. Stayner, C. K., Cunliffe, H. E., Ward, T. A. & Eccles, M. R. Cloning and characterization of the human PAX2 promoter. *J. Biol. Chem.* **273**, 25472–25479 (1998).
20. Shyamala, G. & Guio, M. C. Activation of κ B-specific proteins by estradiol. *Proc. Natl Acad. Sci. USA* **89**, 10628–10632 (1992).
21. Galien, R., Evans, H. F. & Garcia, T. Involvement of CCAAT/enhancer-binding protein and nuclear factor- κ B binding sites in interleukin-6 promoter inhibition by estrogens. *Mol. Endocrinol.* **10**, 713–722 (1996).
22. Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843–852 (2000).
23. Zhang, H. *et al.* Differential gene regulation by the SRC family of coactivators. *Genes Dev.* **18**, 1753–1765 (2004).
24. Cho, K. S., Elizondo, L. I. & Boerkoel, C. F. Advances in chromatin remodeling and human disease. *Curr. Opin. Genet. Dev.* **14**, 308–315 (2004).
25. Watanabe, T. *et al.* Agonistic effect of tamoxifen is dependent on cell type, ERE-promoter context, and estrogen receptor subtype: functional difference between estrogen receptors α and β . *Biochem. Biophys. Res. Commun.* **236**, 140–145 (1997).
26. Robertson, J. A., Bhattacharyya, S. & Ing, N. H. Tamoxifen up-regulates oestrogen receptor- α , c-fos and glyceraldehyde 3-phosphate-dehydrogenase mRNAs in ovine endometrium. *J. Steroid Biochem. Mol. Biol.* **67**, 285–292 (1998).
27. Jones, P. S., Parrott, E. & White, I. N. Activation of transcription by estrogen receptor α and β is cell type- and promoter-dependent. *J. Biol. Chem.* **274**, 32008–32014 (1999).
28. Russo, L. A., Calabro, S. P., Filler, T. A., Carey, D. J. & Gardner, R. M. *In vivo* regulation of syndecan-3 expression in the rat uterus by 17 β -estradiol. *J. Biol. Chem.* **276**, 686–692 (2001).
29. Hague, S. *et al.* Tamoxifen induction of angiogenic factor expression in endometrium. *Br. J. Cancer* **86**, 761–767 (2002).
30. Castro-Rivera, E. & Safe, S. 17 β -estradiol- and 4-hydroxytamoxifen-induced transactivation in breast, endometrial and liver cancer cells is dependent on ER-subtype, cell and promoter context. *J. Steroid Biochem. Mol. Biol.* **84**, 23–31 (2003).
31. Bramlett, K. S. & Burris, T. P. Target specificity of selective estrogen receptor modulators within human endometrial cancer cells. *J. Steroid Biochem. Mol. Biol.* **86**, 27–34 (2003).
32. Paech, K. *et al.* Differential ligand activation of estrogen receptors ER α and ER β at AP1 sites. *Science* **277**, 1508–1510 (1997).
33. Webb, P., Nguyen, P. & Kushner, P. J. Differential SERM effects on corepressor binding dictate ER α activity *in vivo*. *J. Biol. Chem.* **278**, 6912–6920 (2003).
34. Klotz, D. M., Hewitt, S. C., Korach, K. S. & Diaugustine, R. P. Activation of a uterine insulin-like growth factor I signalling pathway by clinical and environmental estrogens: requirement of estrogen receptor- α . *Endocrinology* **141**, 3430–3439 (2000).
35. Jepsen, K. *et al.* Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell* **102**, 753–763 (2000).
36. Katzenellenbogen, B. S. *et al.* Antiestrogens: mechanisms and actions in target cells. *J. Steroid Biochem. Mol. Biol.* **53**, 387–393 (1995).

37. Montano, M. M., Muller, V., Trobaugh, A. & Katzenellenbogen, B. S. The carboxy-terminal F domain of the human estrogen receptor: role in the transcriptional activity of the receptor and the effectiveness of antiestrogens as estrogen antagonists. *Mol. Endocrinol.* **9**, 814–825 (1995).
38. Maulbecker, C. C. & Gruss, P. The oncogenic potential of Pax genes. *EMBO J.* **12**, 2361–2367 (1993).
39. Silberstein, G. B., Dressler, G. R. & Van Horn, K. Expression of the PAX2 oncogene in human breast cancer and its role in progesterone-dependent mammary growth. *Oncogene* **21**, 1009–1016 (2002).
40. Khoubehi, B. *et al.* Expression of the developmental and oncogenic PAX2 gene in human prostate cancer. *J. Urol.* **165**, 2115–2120 (2001).
41. Dressler, G. R. & Douglass, E. C. Pax-2 is a DNA-binding protein expressed in embryonic kidney and Wilms tumour. *Proc. Natl Acad. Sci. USA* **89**, 1179–1183 (1992).
42. Muratovska, A., Zhou, C., He, S., Goodyer, P. & Eccles, M. R. Paired-box genes are frequently expressed in cancer and often required for cancer cell survival. *Oncogene* **22**, 7989–7997 (2003).
43. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89–92 (1983).
44. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nature Rev. Cancer* **4**, 143–153 (2004).
45. Gupta, A., Godwin, A. K., Vanderveer, L., Lu, A. & Liu, J. Hypomethylation of the synuclein γ gene CpG island promotes its aberrant expression in breast carcinoma and ovarian carcinoma. *Cancer Res.* **63**, 664–673 (2003).
46. Nishigaki, M. *et al.* Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. *Cancer Res.* **65**, 2115–2124 (2005).
47. Sato, N. *et al.* Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma. *Cancer Res.* **63**, 4158–4166 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Green for editorial assistance. This work was supported by grants from the National Natural Science Foundation of China and from the '863 Program' and the '973 Program' of the Ministry of Science and Technology of China (to Y.S.).

Author Information The microarray data are deposited in the Gene Expression Omnibus under accession number GSE3013. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.S. (jason@bjmu.edu.cn).

LETTERS

The afterglow and elliptical host galaxy of the short γ -ray burst GRB 050724

E. Berger¹, P. A. Price², S. B. Cenko³, A. Gal-Yam⁴, A. M. Soderberg⁴, M. Kasliwal⁴, D. C. Leonard⁴, P. B. Cameron⁴, D. A. Frail⁵, S. R. Kulkarni⁴, D. C. Murphy¹, W. Krzeminski⁶, T. Piran⁷, B. L. Lee⁸, K. C. Roth⁹, D.-S. Moon³, D. B. Fox⁴, F. A. Harrison³, S. E. Persson¹, B. P. Schmidt¹⁰, B. E. Penprase¹¹, J. Rich¹⁰, B. A. Peterson¹⁰ & L. L. Cowie²

Despite a rich phenomenology, γ -ray bursts (GRBs) are divided¹ into two classes based on their duration and spectral hardness—the long-soft and the short-hard bursts. The discovery of afterglow emission from long GRBs was a watershed event, pinpointing² their origin to star-forming galaxies, and hence the death of massive stars, and indicating³ an energy release of about 10^{51} erg. While theoretical arguments⁴ suggest that short GRBs are produced in the coalescence of binary compact objects (neutron stars or black holes), the progenitors, energetics and environments of these events remain elusive despite recent^{5–8} localizations. Here we report the discovery of the first radio afterglow from the short burst GRB 050724, which unambiguously associates it with an elliptical galaxy at a redshift⁹ $z = 0.257$. We show that the burst is powered by the same relativistic fireball mechanism as long GRBs, with the ejecta possibly collimated in jets, but that the total energy release is 10–1,000 times smaller. More importantly, the nature of the host galaxy demonstrates that short GRBs arise from an old (>1 Gyr) stellar population, strengthening earlier suggestions^{5,6} and providing support for coalescing compact object binaries as the progenitors.

On receipt of the Swift X-ray localization¹⁰ of the short-hard burst GRB 050724 (duration, $T_{90} = 3 \pm 1$ s dominated by an initial spike of 0.25 s; hardness ratio, $F(50\text{--}100\text{ keV})/F(25\text{--}50\text{ keV}) = 0.9 \pm 0.1$; ref. 11) we initiated observations in the radio, near-infrared (NIR) and optical bands at the Very Large Array, the Baade 6.5-m Magellan telescope, and the Swope 40-inch telescope (see Table 1). Within the overall uncertainty of the X-ray position we discovered a point-like radio source and confirmed that it is the radio afterglow by its subsequent variability. Contemporaneous digital image subtraction of our optical and NIR frames from the first and second nights revealed a variable source coincident with the radio afterglow, which we identify as the optical afterglow; this was subsequently confirmed¹² elsewhere. The afterglow is coincident with a bright galaxy, identified¹³ in earlier optical imaging, demonstrating that it is the host galaxy; see Fig. 1.

Motivated by this association, we used the Gemini Multi-Object Spectrograph on the Gemini North telescope to obtain a spectrum of the host galaxy from which we measure a redshift, $z = 0.257$, confirming other measurements⁹; see Fig. 2. At this redshift the fluence¹⁰ of the burst, $F_{\gamma} \approx 6.3 \times 10^{-7} \text{ erg cm}^{-2}$ (15–350 keV), translates to an isotropic-equivalent energy release, $E_{\gamma, \text{iso}} \approx 4 \times 10^{50} \text{ erg}$; this includes a bolometric correction of a factor of four, the average

ratio of the 20–2,000 keV to the 25–300 keV luminosity in the BATSE short burst sample. The X-ray luminosity of the afterglow at $t = 10$ h, a proxy¹⁴ for the kinetic energy of the blast wave, is $L_{X, \text{iso}} \approx 4.2 \times 10^{44} \text{ erg s}^{-1}$. Both of these quantities are at least an order of magnitude lower^{3,15} than for the long GRBs.

The minimum initial Lorentz factor of the ejecta is between 80 and 160, based¹⁰ on a peak flux, $f_p \approx 3.9 \text{ photons cm}^{-2} \text{ s}^{-1}$, and the duration of the initial hard spike¹¹ of 0.25 s. This large value, similar to those inferred¹⁶ for long GRBs, indicates a relativistically expanding fireball, which in turn produces the afterglow emission. GRB 050724 is the first short burst with radio, optical/NIR and X-ray afterglows, and we are therefore in a unique position to derive the properties of the fireball and burst environment. Using a standard synchrotron power-law spectrum¹⁷ fit to the afterglow data at

Table 1 | Afterglow observations of GRB 050724 in the radio, optical, and near-infrared bands

Epoch (UT)	Δt (h)	Telescope	Band	Flux (μJy)
Jul 25.01	11.6	Magellan/PANIC	K	38.7 ± 1.4
Jul 25.98	34.9	Magellan/PANIC	K	< 4.6
Jul 25.02	12.0	Swope 40-inch	I	8.4 ± 0.2
Jul 25.11	14.2	Swope 40-inch	I	11.1 ± 0.9
Jul 26.05	36.7	Swope 40-inch	I	< 4.0
Jul 25.09	13.7	VLA	8.46	173 ± 30
Jul 26.21	40.5	VLA	8.46	465 ± 29

For the radio observations we list the observing band in GHz, while for the optical and NIR data we list the filter. In all Very Large Array (VLA) observations we used the extra-galactic sources 3C 286 and J1626 – 298 for flux and phase calibration, respectively. The data were reduced and analysed using the Astronomical Image Processing System, and the flux density and uncertainty were measured from the resulting maps. The NIR data were obtained with Persson's Auxiliary Nasmyth Infrared Camera and consisted of sixty-six 20-s images at each epoch. The individual images were processed in the standard manner and corrected for distortion. Astrometry was performed relative to nine 2-Micron All-Sky Survey (2MASS) sources resulting in an astrometric accuracy of about $0.1''$. The optical observations consisted of two 900-s images in the first two epochs and three 900-s images in the third epoch. The data were processed in the standard manner, and astrometry was performed relative to 65 USNO (United States Naval Observatory) stars, resulting in a root-mean-square (r.m.s.) uncertainty of about $0.15''$. Photometry of the NIR afterglow was performed using the 'NN2' method²⁷. Errors in the subtracted fluxes were measured from the r.m.s. deviation of fluxes within apertures randomly distributed over the background of the subtracted images. Flux calibration was performed relative to sources from the 2MASS catalogue; our resulting absolute calibration is limited by statistical errors in the faint catalogue sources to an accuracy of about 5%. Photometry of the optical afterglow was performed using image subtraction, and the flux measurements carry a systematic uncertainty of about 0.15 mag, which affects all epochs in the same manner. Therefore, the observed re-brightening between 12 and 14.2 h is significant at about 3σ . The optical and NIR fluxes are not corrected for Galactic extinction (see Fig. 3).

¹Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA. ²Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. ³Space Radiation Laboratory 220-47, ⁴Caltech Optical Observatories 105-24, California Institute of Technology, Pasadena, California, 91125 USA. ⁵National Radio Astronomy Observatory, PO Box 0, Socorro, New Mexico 87801, USA. ⁶Las Campanas Observatory, Carnegie Observatories, Casilla 601, La Serena, Chile. ⁷Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel. ⁸Department of Astronomy and Astrophysics, University of Toronto, Toronto, Ontario M5S 3H8, Canada. ⁹Gemini Observatory, 670 N. Aohoku Place Hilo, Hawaii 96720, USA. ¹⁰RSAA, ANU, Mt Stromlo Observatory, via Cotter Rd, Weston Creek, ACT 2611, Australia. ¹¹Pomona College Department of Physics and Astronomy, 610 N. College Ave, Claremont, California 91711, USA.

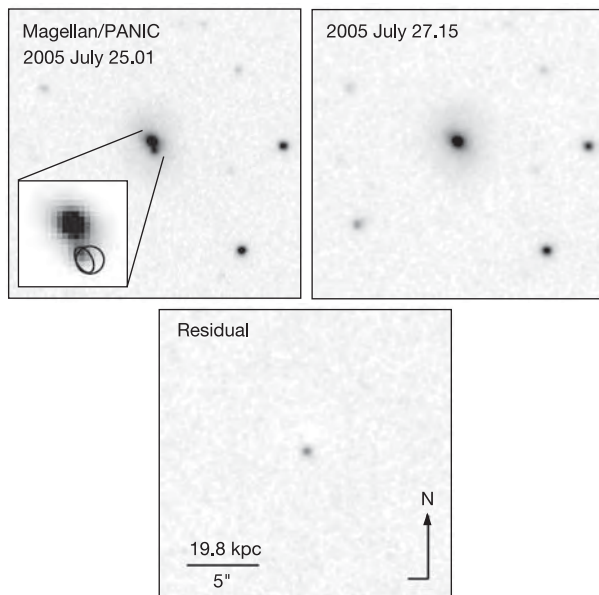


Figure 1 | Near-infrared K-band images of the afterglow and host galaxy of GRB 050724. The afterglow has completely faded between 11.6 and 34.9 hours after the burst, indicating a flux decay rate of $\alpha < -1.9$ ($F_p \propto t^\alpha$). The position of the NIR afterglow is $\alpha = 16\text{ h } 24\text{ min } 44.38\text{ s}$, $\delta = -27^\circ 32' 27.5''$, with an uncertainty of about $0.1''$ in each coordinate. The inset shows the VLA radio position (ellipse; $\alpha = 16\text{ h } 24\text{ min } 44.37\text{ s}$, $\delta = -27^\circ 32' 27.5''$) and the Chandra X-ray position²⁸ (circle; $\alpha = 16\text{ h } 24\text{ min } 44.36\text{ s}$, $\delta = -27^\circ 32' 27.5''$). These positions are fully consistent within the measurement uncertainty: the radio–NIR offset is $\Delta\alpha = 0.12 \pm 0.11''$, $\Delta\delta = 0.19 \pm 0.21''$, while the X-ray–NIR offset is $\Delta\alpha = 0.28 \pm 0.22''$ and $\Delta\delta = 0.01 \pm 0.22''$. The host galaxy brightness corrected for Galactic extinction is $K = 14.88 \pm 0.03\text{ mag}$ ($3.5''$ aperture). Using standard cosmological parameters ($\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $H_0 = 71\text{ km s}^{-1}\text{ Mpc}^{-1}$) the absolute magnitude is $M_K = -24.7\text{ (mag)}$, including a k -correction factor of 1.05 mag. This suggests that the host is a bright galaxy with a luminosity, $L \approx 1.6 L^*$ by comparison to the luminosity function derived²⁹ from the 2dF Galaxy Redshift Survey and 2MASS; see also ref. 9. The host magnitude in the optical I-band is $I = 18.63 \pm 0.2\text{ mag}$, indicating a red $I - K \approx 2.56 \pm 0.2\text{ mag}$. The radial surface brightness distribution, measured using elliptical isophotes, follows a de Vaucouleurs $r^{1/4}$ profile with an effective radius of 6 kpc and a central surface brightness of $\mu_K \approx 19.5\text{ mag arcsec}^{-2}$. The ellipticity of the host is about 0.17, indicating an E2 Hubble classification.

$t = 12\text{ h}$ we find an isotropic-equivalent kinetic energy of $E_{K,\text{iso}} \approx 1.5 \times 10^{51}\text{ erg}$, a density of $n \approx 0.1\text{ cm}^{-3}$, and fractions of energy in the relativistic electrons and magnetic field of $\epsilon_e \approx 0.04$ and $\epsilon_B \approx 0.02$, respectively; a mild degeneracy between n and $E_{K,\text{iso}}$ marginally accommodates a density as low as 0.02 cm^{-3} and an energy as high as $3 \times 10^{51}\text{ erg}$ (see Fig. 3). A comparison of $E_{\gamma,\text{iso}}$ and $E_{K,\text{iso}}$ indicates a radiative efficiency of about 20%, similar to the long GRBs.

The fading rate of the NIR afterglow emission between 12 and 35 hours after the burst is steeper than $F_p \propto t^{-1.9}$ (Table 1). This is suggestive of a collimated explosion, or jet¹⁸. The flat or rising optical light curve between 12 and 14.2 h, suggests that the jet break time is $\sim 1\text{ d}$, corresponding to an opening angle³, $\theta_j \approx 0.15\text{ rad}$ (for $n = 0.1\text{ cm}^{-3}$). In this framework the true energy release is $E_\gamma \approx 4 \times 10^{48}\text{ erg}$ and $E_K \approx 1.7 \times 10^{49}\text{ erg}$, two orders of magnitude below the energy release of long GRBs. We note that jet breaks are achromatic, and our tentative claim can be tested with additional radio data. We conclude from this discussion that GRB 050724 is powered by the same fireball mechanism as long GRBs, with similar micro-physical properties. The fundamental difference is that the total energy release is a few orders of magnitude smaller than in long GRBs.

We now turn to the nature of the progenitor system, as revealed

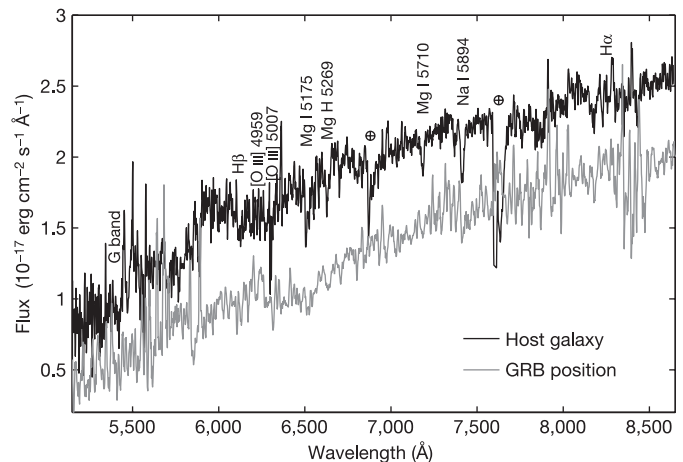


Figure 2 | Optical spectra of the host galaxy of GRB 050724. The black line shows a spectrum taken through the centre of the host galaxy with the Gemini Multi-Object Spectrograph on the Gemini North telescope on 2005 July 27 UT. The observations consisted of $4 \times 1,800\text{ s}$ integrations through a $1''$ wide slit. The spectrum was processed with the standard gmos reduction tasks in IRAF (Image Reduction and Analysis Facility) before subtracting the sky, and extracting the spectra. Flux calibration was achieved through an archival observation of a spectrophotometric standard star, taken with a slightly different instrumental set-up; as such, the flux calibration is not completely accurate, but is indicative. We detect the Na D lines in absorption at a mean redshift of $z = 0.257 \pm 0.001$, confirming another measurement⁹. We also detect absorption lines corresponding to Mg b (5174), MgH (5269) and Mg I (5710). We place a 3σ limit of about $1.5 \times 10^{-17}\text{ erg cm}^{-2}\text{ s}^{-1}$ on the flux of H α , corresponding at the redshift of the host to a limit of $< 0.02 M_\odot\text{ yr}^{-1}$. The lack of a prominent H β absorption feature indicates¹⁹ a stellar population older than $\sim 1\text{ Gyr}$. The grey line is a spectrum obtained at the position of GRB 050724 with the Low-Resolution Imaging Spectrometer on the Keck II telescope on 2005 July 28 UT. A slit position angle of 90° was chosen to minimize contribution from the galaxy light. This allows us to place constraints on any residual star formation at the position of the GRB. The observation consisted of a single 2,240 s exposure with a $1.5''$ wide slit. Flux calibration was performed using the spectrophotometric standard BD + $17^\circ 47' 08''$ (red) and BD + $28^\circ 42' 11''$ (blue). The shape of the spectrum is not completely accurate because of slit losses, but the lack of detectable H α emission allows us to place a limit of $0.03 M_\odot\text{ yr}^{-1}$ on the star-formation rate at the position of the GRB.

from the properties of the host galaxy. The spectrum indicates that the host is an early-type galaxy⁹, with a stellar population that is older than $\sim 1\text{ Gyr}$, based on the lack of detectable Balmer H β absorption¹⁹. From the limit on H α emission we find that the overall star-formation rate is $< 0.02 M_\odot\text{ yr}^{-1}$, and more importantly, at the location of GRB 050724 we place a limit of $< 0.03 M_\odot\text{ yr}^{-1}$ (Fig. 2). The red colour and surface brightness profile of the galaxy indicate an elliptical (E2) classification. The position of GRB 050724 is offset by about $2,570 \pm 80\text{ pc}$ from the centre of the host, corresponding to $0.4r_e$, where $r_e \approx 6\text{ kpc}$ is the host galaxy's effective radius. This offset is smaller²⁰ than for 80% of the long bursts.

The association of the burst with an elliptical galaxy dominated by an old stellar population is unlike any of the long GRBs localized to date, which invariably occur²¹ in star-forming galaxies. This leads us to conclude that the progenitor of GRB 050724 was not a massive star, but was instead related to an old stellar population. Theoretical considerations suggest⁴ that short bursts arise from the coalescence of binary systems undergoing angular momentum loss via gravitational radiation. Both neutron star–neutron star and neutron star–black hole systems have been proposed, leading to a prediction²² of a wide distribution of coalescence timescales ($\sim 10^7 - 10^{10}\text{ yr}$) and hence offsets (few to 10^3 kpc). For delayed mergers, the median redshift is predicted²³ to be lower compared to the bulk of the star-formation activity, that is $z \ll 1$.

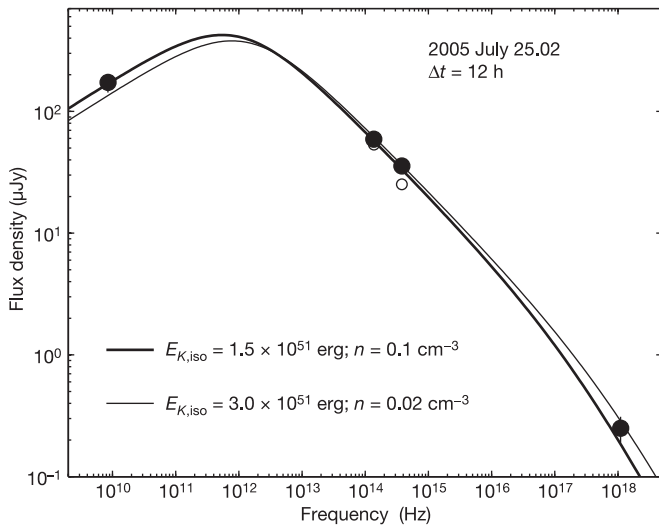


Figure 3 | Radio to X-ray spectral energy distribution of the afterglow emission 12 h after the burst. In the optical and NIR bands the open circles are the measured fluxes without a correction for Galactic extinction. We find that the Galactic extinction along the line of sight required for reconciling the optical, NIR and X-ray fluxes is $A_V \approx 2.7$ mag, about 35% higher than the tabulated value³⁰. The inferred extinction is in very good agreement with the increased hydrogen column density, $N_H \approx 5.6 \times 10^{21} \text{ cm}^{-2}$, inferred¹¹ from the X-ray afterglow, and it indicates that the excess absorption has a Galactic, rather than host galaxy, origin. The lines are synchrotron models¹⁷ of the afterglow emission. We find a best-fit solution with an energy of $E_{K,iso} \approx 1.5 \times 10^{51}$ erg, a density of $n \approx 0.1 \text{ cm}^{-3}$, and fractions of energy in the relativistic electrons and magnetic field of $\epsilon_e \approx 0.04$ and $\epsilon_B \approx 0.02$, respectively. A slight degeneracy between the energy and density is shown by the thin line, which marginally fits the data.

The old age of the host's stellar population, the lack of detectable current star formation at the position of the burst, and the low redshift compared to most long GRBs point to a binary progenitor that had a coalescence time of ≥ 1 Gyr. The small offset, however, suggests that the kick velocity imparted to the system was probably too small to unbind it from the host. In this context a comparison to two recent short GRBs is illustrative. GRB 050509b was possibly associated^{5,6} with an elliptical galaxy at $z = 0.225$, but the poor localization ($9.3''$ radius⁶) also allowed an association with higher-redshift star-forming galaxies. The results on GRB 050724 now support the claimed association with the elliptical galaxy. On the other hand, GRB 050709 was precisely localized within 3 kpc of a star-forming galaxy²⁴ at $z = 0.16$. While this association does not allow a definitive argument against a massive star origin, it suggests that the progenitors of short GRBs occur in diverse environments, and with a range of coalescence timescales. This scenario is similar to that of type Ia supernovae²⁵.

We conclude with the following intriguing possibilities. First, the isotropic-equivalent prompt energy release appears to correlate with the burst duration, such that the luminosity is nearly constant, $L_{\gamma,iso} \approx (3\text{--}15) \times 10^{50}$ erg. This may explain why the afterglow of the 40-ms duration GRB 050509b was significantly fainter than those of GRBs 050709 and 050724. Second, the small offsets and low redshifts of the latter two bursts may contrast with population synthesis models, which predict^{22,23} 90% of the offsets to be >10 kpc, and a median redshift $z \approx 0.5\text{--}1$. Finally, if the beaming inferred for GRB 050724 is typical of the short burst population, then this implies that the true event rate of short bursts is about fifty times higher than observed. This suggests that $\lesssim 10\%$ of neutron star–neutron star and neutron star–black hole binaries²⁶ end their lives in GRB explosions. The continued detection of short GRBs by Swift will

allow these conclusions to be tested through the distribution of energies, jet angles and offsets.

Received 3 August; accepted 14 September 2005.

- Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
- Pian, E. *et al.* Hubble Space Telescope Imaging of the optical transient associated with GRB 970508. *Astrophys. J.* **492**, L103–L106 (1998).
- Frail, D. A. *et al.* Beaming in gamma-ray bursts: Evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
- Katz, J. I. & Canel, L. M. The long and the short of gamma-ray bursts. *Astrophys. J.* **471**, 915–920 (1996).
- Bloom, J. S. *et al.* Closing in on a short-hard burst progenitor: Constraints from early-time optical imaging and spectroscopy of a possible host galaxy of GRB 050509b. Preprint at (<http://arXiv.org/astro-ph/0505480>) (2005).
- Gehrels, N. *et al.* A short γ -ray burst apparently associated with an elliptical galaxy at redshift $z = 0.225$. *Nature* **437**, 851–854 (2005).
- Fox, D. B. *et al.* The afterglow of GRB 050709 and the nature of the short-hard γ -ray bursts. *Nature* **437**, 845–850 (2005).
- Hjorth, J. *et al.* The optical afterglow of the short γ -ray burst GRB 050709. *Nature* **437**, 859–861 (2005).
- Prochaska, J. X. *et al.* GRB 050724: secure host redshift from Keck. *GCN Circ.* **3700** (2005).
- Krimm, H. *et al.* GRB050724: refined analysis of the Swift-BAT possible short bursts. *GCN Circ.* **3667** (2005).
- Barthelmy, S. D. *et al.* An origin for short γ -ray bursts unassociated with current star formation. *Nature* doi:10.1038/nature04392 (this issue).
- D'Avanzo, P. *et al.* GRB050724: VLT observations of the variable source. *GCN Circ.* **3690** (2005).
- Bloom, J. S., Dupree, A., Chen, H.-W. & Prochaska, J. X. GRB050724: GMOS imaging and spectroscopy. *GCN Circ.* **3679** (2005).
- Freedman, D. L. & Waxman, E. On the energy of gamma-ray bursts. *Astrophys. J.* **547**, 922–928 (2001).
- Berger, E., Kulkarni, S. R. & Frail, D. A. A standard kinetic energy reservoir in gamma-ray burst afterglows. *Astrophys. J.* **590**, 379–385 (2003).
- Lithwick, Y. & Sari, R. Lower limits on Lorentz factors in gamma-ray bursts. *Astrophys. J.* **555**, 540–545 (2001).
- Granot, J. & Sari, R. The shape of spectral breaks in gamma-ray burst afterglows. *Astrophys. J.* **568**, 820–829 (2002).
- Sari, R., Piran, T. & Halpern, J. P. Jets in gamma-ray bursts. *Astrophys. J.* **519**, L17–L20 (1999).
- Dressler, A. & Gunn, J. E. Spectroscopy of galaxies in distant clusters. II—The population of the 3C 295 cluster. *Astrophys. J.* **270**, 7–19 (1983).
- Bloom, J. S., Kulkarni, S. R. & Djorgovski, S. G. The observed offset distribution of gamma-ray bursts from their host galaxies: a robust clue to the nature of the progenitors. *Astron. J.* **123**, 1111–1148 (2002).
- Christensen, L., Hjorth, J. & Gorosabel, J. UV star-formation rates of GRB host galaxies. *Astron. Astrophys.* **425**, 913–926 (2004).
- Fryer, C. L., Woosley, S. E. & Hartmann, D. H. Formation rates of black hole accretion disk gamma-ray bursts. *Astrophys. J.* **526**, 152–177 (1999).
- Guetta, D. & Piran, T. The luminosity and redshift distributions of short-duration GRBs. *Astron. Astrophys.* **435**, 421–426 (2005).
- Price, P. A. GRB 050709: spectroscopy. *GCN Circ.* **3605** (2005).
- van den Bergh, S. The frequency of SN Ia in galaxies of different Hubble type. *Publ. Astron. Soc. Pacif.* **102**, 1318–1320 (1990).
- Phinney, E. S. The rate of neutron star binary mergers in the universe—Minimal predictions for gravity wave detectors. *Astrophys. J.* **380**, L17–L21 (1991).
- Barris, B. J., Tonry, J. L., Novicki, M. C. & Wood-Vasey, W. M. The NN2 flux difference method for constructing variable object light curves. Preprint at (<http://arXiv.org/astro-ph/0507584>) (2005).
- Burrows, D. *et al.* GRB 050724: Chandra observations of the X-ray afterglow. *GCN Circ.* **3697** (2005).
- Cole, S. *et al.* The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions. *Mon. Not. R. Astron. Soc.* **326**, 255–273 (2001).
- Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).

Acknowledgements We are, as always, indebted to S. Barthelmy and the GCN. GRB research at Carnegie and Caltech is supported in part by funds from NASA. E.B. and A.G.Y. are supported by NASA through Hubble Fellowship grants awarded by the Space Telescope Science Institute, which is operated by AURA, Inc., for NASA. The VLA is operated by the National Radio Astronomy Observatory, a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.B. (eberger@ociw.edu).

An origin in the local Universe for some short γ -ray bursts

N. R. Tanvir¹, R. Chapman¹, A. J. Levan¹ & R. S. Priddey¹

Gamma-ray bursts (GRBs) divide into two classes¹: ‘long’, which typically have initial durations of $T_{90} > 2$ s, and ‘short’, with durations of $T_{90} < 2$ s (where T_{90} is the time to detect 90% of the observed fluence). Long bursts, which on average have softer γ -ray spectra², are known to be associated with stellar core-collapse events—in some cases simultaneously producing powerful type Ic supernovae^{3–5}. In contrast, the origin of short bursts has remained mysterious until recently. A subsecond intense ‘spike’ of γ -rays during a giant flare from the Galactic soft γ -ray repeater, SGR 1806–20, reopened an old debate over whether some short GRBs could be similar events seen in galaxies out to ~ 70 Mpc (refs 6–10; redshift $z \approx 0.016$). Shortly after that, localizations of a few short GRBs (with optical afterglows detected in two cases^{11,12}) have shown an apparent association with a variety of host galaxies at moderate redshifts^{11–14}. Here we report a correlation between the locations of previously observed short bursts and the positions of galaxies in the local Universe, indicating that between 10 and 25 per cent of short GRBs originate at low redshifts ($z < 0.025$).

The satellite-based γ -ray detector CGRO/BATSE triggered on roughly 500 short-duration bursts during its nine-year lifetime. Unfortunately, the (1σ) positional uncertainties for these bursts were typically many degrees, giving limited information to help identify even their host galaxies, let alone their progenitors. A handful of short GRBs were localized to smaller error boxes by the Interplanetary Network and the Beppo-SAX and HETE-II satellites, but these only showed an absence of bright candidate host galaxies or afterglows^{15,16}. Recently, however, afterglow detections for three short GRBs have associated them with galaxies at redshifts $z \approx 0.2$: the X-ray afterglow of GRB 050509B was close to a bright elliptical galaxy at $z = 0.225$, suggesting physical association^{13,14}; GRB 050709 produced an optical afterglow locating it to a late-type galaxy at $z = 0.16$ (ref. 11); and GRB 050724 exhibited an afterglow located in another elliptical galaxy at $z = 0.257$ (ref. 12). The energetics of these bursts, and their association with a variety of host galaxies (including those with only old stellar populations) provide support for the view that some fraction of short GRBs arise from the coalescence of neutron-star/neutron-star (NS–NS) binaries^{17–19}.

On the other hand, the recent observation of the ‘hypergiant’ flare from SGR 1806–20 (refs 6, 7) re-ignited interest in the idea that some short bursts could be distant SGR flares. Although previous giant flares were bright enough to have been seen by BATSE perhaps as far as the Virgo cluster, the SGR 1806–20 event would have been detectable out to several tens of megaparsecs, appearing very much like a short-hard GRB.

If even a proportion of short bursts originate in nearby galaxies, then despite poor localizations they may show a measurable spatial correlation with the positions of low-redshift galaxies. The PSCz galaxy redshift survey²⁰ makes an appropriate comparison data set for the all-sky BATSE data because, being IRAS-selected, it suffers less

from incompleteness at low Galactic latitudes than other nearby redshift surveys (although other catalogues show similar results, as described in Supplementary Information).

We considered all 400 $T_{90} < 2$ s bursts with localizations better than 10 degrees from the BATSE catalogue (4B(R); ref. 21 together with web supplement cited therein). A cross-correlation plot between these GRBs and the (1,070) PSCz galaxies with heliocentric recession velocity, $v_r \leq 2,000$ km s^{−1} is shown in Fig. 1. This sample includes most galaxies within about 25 Mpc, encompassing the local supercluster, and specifically the Virgo, Fornax and Ursa Major clusters. A clear positive correlation is revealed, which is even stronger when the galaxies are restricted to earlier morphological types (specifically Sbc and earlier; that is, T-type ≤ 4). The figure also shows, as expected, that the long-duration bursts are uncorrelated with these galaxies.

However, the cross-correlation function is not ideal for this

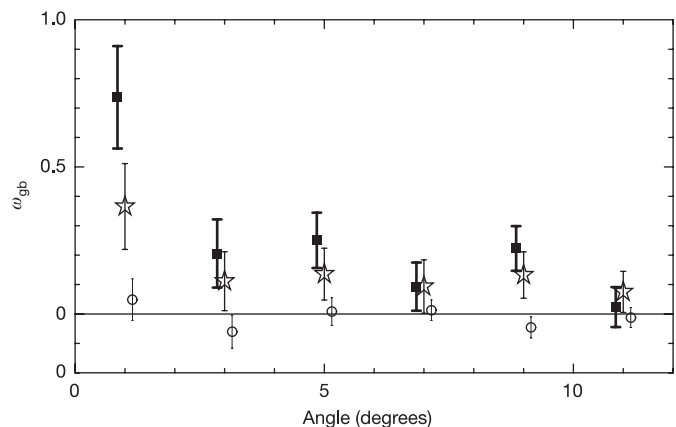


Figure 1 | Angular cross-correlation functions for BATSE short bursts with nearby galaxies from the PSCz catalogue. Two-point angular correlation functions, ω_{gb} (in two-degree bins), are shown for 400 BATSE short bursts ($T_{90} < 2$ s and statistical position uncertainty ≤ 10 degrees) with nearby galaxies from the PSCz catalogue (recession velocities $v_r \leq 2,000$ km s^{−1} ≈ 28 Mpc distant). Filled circles show the correlation with all galaxy types (1,072 galaxies), and bold square symbols represent the same function but with galaxies restricted to earlier morphological types (T-type ≤ 4 , Sbc and earlier, 709 galaxies). Open circles show the same function for long-duration bursts compared with all galaxy types. Points from the different functions have been offset slightly in angle for clarity. The short bursts exhibit clear ($>3\sigma$) correlation at low angles with the earlier-type galaxies, and a $>2\sigma$ correlation with all galaxy types. In contrast, as would be expected, the long bursts (1,481 events) show no measurable correlation with local galaxies. The highly anisotropic distribution on the sky of galaxies within this region makes it possible to detect this correlation signal. Uncertainties (error bars represent 1σ) in each bin were determined from Monte Carlo simulations, but true errors are not independent from bin to bin.

¹Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK.

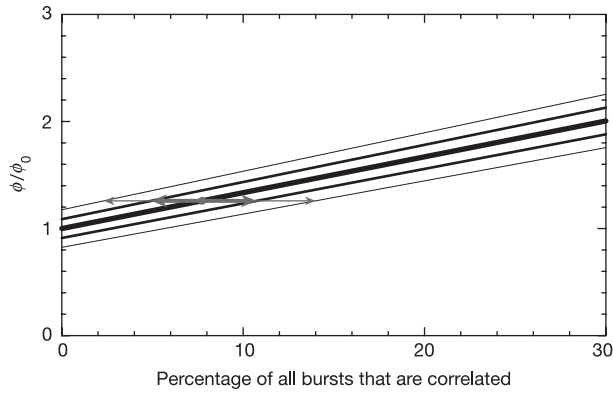


Figure 2 | Proportion of short bursts correlated with galaxies within $v \leq 2,000 \text{ km s}^{-1}$. This figure summarizes an analysis of many simulated BATSE short-burst samples in which a proportion of bursts are laid down correlated with the PSCz galaxy positions and the remainder are placed at random. In this case, the galaxy sample is restricted to morphological T-type ≤ 4 and within recession velocity $v = 2,000 \text{ km s}^{-1}$. The bold diagonal line represents the value of Φ/Φ_0 (see main text) as a function of the proportion of bursts in the simulated sample whose positions were seeded by the galaxy positions (random errors based on those in the BATSE catalogue were used to find offsets from these seed positions). The thinner lines show the 1σ and 2σ deviations around this line according to the simulations. The level of Φ/Φ_0 for the real data is illustrated by the horizontal arrow, which spans the 2σ (small arrow) and 1σ (large arrow) ranges. Thus the possibility of no correlation is rejected at more than the 3σ level, and a correlated fraction around $\sim 8\%$ is indicated (95% confidence limits $\sim 2\%$ to $\sim 14\%$). As discussed in the text, there are reasons for expecting this to be an underestimate. Note that using all T-types produces a consistent, but less well constrained result.

purpose as errors in different bins are not independent, and it makes no direct use of known BATSE instrumental characteristics. In order to optimize the search for a correlation signal, we compute Φ , the sum of all burst-galaxy pairs weighted by the probability that they could be seen at the observed separation (or greater) if they were truly associated, and further weighted inversely by the burst error-circle size:

$$\Phi = \sum_i^{\text{All bursts}} \sum_j^{\text{All galaxies}} \frac{1}{\varepsilon_i} \int_{\theta_{ij}}^{\infty} \frac{1}{\sqrt{2\pi}\varepsilon_i} \exp\left(\frac{-\theta^2}{2\varepsilon_i^2}\right) d\theta$$

where θ_{ij} is the separation between the i th burst and the j th galaxy, and ε_i is the error circle (statistical and systematic²²) of the i th burst position. Further discussion of the BATSE instrumental characteristics²³ is provided in Supplementary Information.

To quantify the significance of the observed value of Φ , we also compute Φ_0 , which is the mean of a large number of simulated random burst distributions (each with the same number of positions as the number of bursts under consideration, the same positional errors, and distributed on the sky according to the known BATSE sky exposure map²⁴) correlated against the same set of (T-type ≤ 4) PSCz galaxy positions. The spread of simulated results around Φ_0 allows us to test the null hypothesis that there is no correlation between the positions of bursts and galaxies. This null hypothesis is rejected at the 99.9% level, confirming the indications of the cross-correlation function.

Next we attempt to estimate the proportion of bursts that are associated with nearby galaxies. To do this, we constructed many more artificial short-burst data sets, this time with both a random component and a component correlated with the galaxies from the PSCz galaxy catalogue with $v \leq 2,000 \text{ km s}^{-1}$ and T-type ≤ 4 (that is, with a 'host' selected at random from the catalogue and with positions smeared according to the real error circles from BATSE). Figure 2 shows that the observed signal could be explained by a

correlated component representing between 6% and 12% of BATSE bursts (1σ range). In fact, this is likely to be an underestimate, as our catalogue certainly does not contain all the galaxies in this volume, just the infrared-bright ones, and there remains a thin zone of avoidance around the Galactic plane, and a region unsurveyed by IRAS that the PSCz survey does not cover (amounting to about 16% of the sky).

Repeating this procedure in bins of recession velocity $v = 2,000$ – $5,000 \text{ km s}^{-1}$ and $v = 5,000$ – $8,000 \text{ km s}^{-1}$, we continue to find significant correlations (at nearly the 2σ level in each bin). This provides strong confirmation that the correlation seen in the low-redshift bin was not simply a chance coincidence. The cumulative proportion of correlated bursts as a function of limiting cut-off velocity is shown in Fig. 3. Inevitably it becomes harder to detect correlations as the galaxy (and presumably detected burst) volume density decreases with increasing distance, and the angular size of large-scale structure projected on the sky also reduces. The range of percentage correlated bursts conservatively suggests that a total proportion of between 10% and 25% of BATSE bursts originate within $\sim 100 \text{ Mpc}$.

Our results are broadly consistent (discussed further in Supplementary Information) with previously reported upper limits for the proportion of BATSE short bursts originating at low redshifts^{6–10}. They also explain the intriguing finding that short-burst positions on the sky show evidence for a weak auto-correlation signal^{25,26}.

At first sight, the most likely explanation of our result is that we are detecting a low-redshift population of short bursts associated with SGRs. However, SGRs being the remnants of short-lived massive stars, it is then rather surprising that a stronger correlation is seen with earlier-type galaxies, which include some galaxies with little recent star formation. Furthermore, although a positive correlation signal is seen when restricting the burst sample to only those with $T_{90} < 0.5 \text{ s}$, a somewhat stronger signal is seen with the $T_{90} > 0.5 \text{ s}$ short bursts, which is again surprising given that the spike from SGR1806–20 had a duration⁶ of around 0.2 s .

An alternative possibility is that some or all of the low-redshift short bursts have the same progenitors as the recently discovered short-burst population at redshifts $z > 0.1$. This would have the merit of simplicity, but as we show in Fig. 4, the higher-redshift bursts so far detected would have been rather brighter than any detected by BATSE if they had occurred at $z < 0.02$. Thus a rather broad intrinsic luminosity function, perhaps comparable to that of

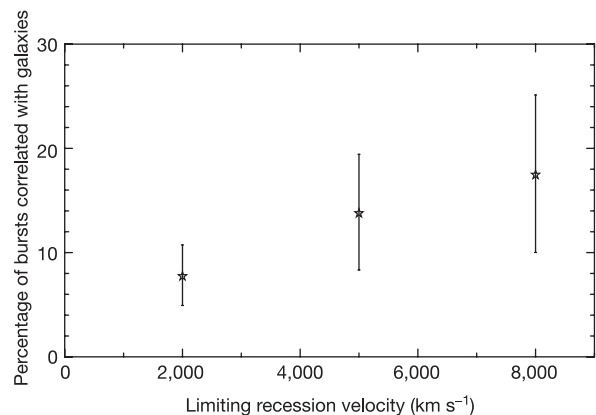


Figure 3 | The percentage of correlated bursts with increasing redshift. This figure shows the results of repeating the analysis of Fig. 2 for two further velocity-limited samples. Both the expected number of correlated bursts and 1σ ranges are shown. We emphasize that these figures are arrived at by comparing the relative values of Φ/Φ_0 for many simulated burst data sets with the real galaxy data set in each case. At higher redshifts the implied proportion of correlated bursts increases, as we would expect, but the statistical constraints also become weaker.

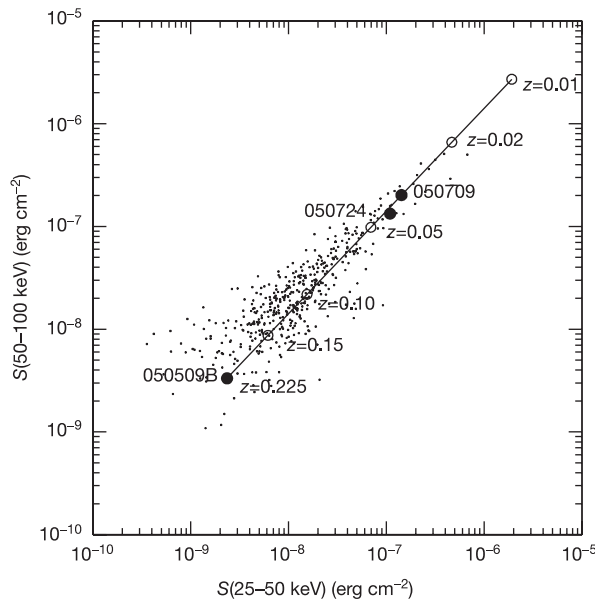


Figure 4 | The positions of GRB 050509b, GRB050709 and GRB050724 (bold points) in the BATSE fluence distribution. The measured γ -ray fluence (S) of GRB 050509b (ref. 14) was very low ($9.5 \times 10^{-9} \text{ erg cm}^{-2}$ in the 15–150 keV band) and, when converted into the full BATSE passband (25–300 keV), is essentially the faintest burst within this catalogue. As illustrated by the open points, if this burst had occurred within 100 Mpc it would lie in the brightest 1% of bursts, although it would be brighter than any BATSE bursts were it closer than 30 Mpc—the volume in which we measure our most significant correlation. The other two recently claimed short-burst identifications are brighter^{29,30}, and therefore if they had occurred at lower redshift would certainly have been the brightest in the BATSE sample. This implies that the bursts responsible for the correlation we measure are most probably intrinsically much less luminous than these three, although, as discussed in the text, a broad luminosity function might accommodate both with the same class of progenitor.

the long-duration bursts, is required to accommodate reasonable numbers of both local and cosmological examples within the BATSE sample. A combination of differing progenitor masses and beaming could plausibly give rise to such a broad luminosity function for NS–NS binary mergers²⁷.

If both cosmological and local short-GRBs arise from NS–NS coalescence, then their association with at least intermediate-age and possibly old stellar populations is to be expected (owing to the inspiral lifetime of such binaries). Furthermore, the rate implied by our work of several bursts per year within 100 Mpc is in good agreement with recent calculations of the rate density of such events, based on the observed double neutron star population in the Milky Way²⁸. The presence of such a large local population of NS–NS merger events would be encouraging for their detection prospects with upcoming gravity wave detectors.

Received 15 June; accepted 6 October 2005.

- Norris, J. P., Cline, T. L., Desai, U. D. & Teegarden, B. J. Frequency of fast narrow gamma-ray bursts. *Nature* **308**, 434–435 (1984).
- Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
- Woosley, S. E. Gamma-ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).

- Hjorth, J. *et al.* A very energetic supernova associated with the gamma-ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
- Stanek, K. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
- Hurley, K. *et al.* An exceptionally bright flare from SGR 1806–20 and the origins of short duration γ -ray bursts. *Nature* **434**, 1098–1103 (2005).
- Palmer, D. *et al.* A giant γ -ray flare from the magnetar SGR 1806–20. *Nature* **434**, 1107–1109 (2005).
- Nakar, E., Gal-Yam, A., Piran, T. & Fox, D. B. The distances to short-hard GRBs and the SGR connection. *Astrophys. J.* (2005)(in the press); preprint at (<http://arxiv.org/astro-ph/0502148>).
- Popov, S. B. & Stern, B. E. Soft gamma repeaters outside the Local group. *Mon. Not. R. Astron. Soc.* (submitted); preprint at (<http://arxiv.org/astro-ph/0503532>) (2005).
- Lazzati, D. A certain flare. *Nature* **434**, 1075–1076 (2005).
- Hjorth, J. *et al.* The optical afterglow of the short γ -ray burst GRB 050709. *Nature* **437**, 859–861 (2005).
- Berger, E. *et al.* The afterglow and elliptical host galaxy of the short γ -ray burst GRB 050724. *Nature* doi:10.1038/nature04238 (in the press); preprint at (<http://arxiv.org/astro-ph/0508115>) (2005).
- Bloom, J. *et al.* Closing in on a short-hard burst progenitor: constraints from early-time optical imaging and spectroscopy of a possible host galaxy of GRB 050509b. *Astrophys. J.* (in the press); preprint at (<http://arxiv.org/astro-ph/0505480>).
- Gehrels, N. *et al.* A short γ -ray burst apparently associated with an elliptical galaxy at redshift $z = 0.225$. *Nature* **437**, 851–854 (2005).
- Klotz, A., Boer, M. & Atteia, J.-L. Observational constraints on the afterglow of GRB 020531. *Astron. Astrophys.* **404**, 815–818 (2003).
- Gorosabel, J. *et al.* Constraints on the afterglow of the short/hard burst GRB 010119. *Astron. Astrophys.* **383**, 112–117 (2002).
- Eichler, D., Livio, M., Piran, T. & Schramm, D. N. Nucleosynthesis, neutrinos and gamma-rays from coalescing neutron stars. *Nature* **340**, 126–128 (1989).
- Mochkovitch, R., Hernanz, M., Isern, J. & Martin, X. Gamma-ray bursts as collimated jets from neutron star/black hole mergers. *Nature* **361**, 236–238 (1993).
- Rosswog, S., Ramirez-Ruiz, E. & Davies, M. B. High resolution calculations of merging neutron stars III: Gamma-ray bursts. *Mon. Not. R. Astron. Soc.* **345**, 1077–1090 (2003).
- Saunders, W. *et al.* The PSCz catalogue. *Mon. Not. R. Astron. Soc.* **317**, 55–63 (2000).
- Paciesas, W. *et al.* The Fourth BATSE Gamma-Ray Burst Catalog (revised). *Astrophys. J. Suppl.* **122**, 465–495 (1999).
- Briggs, M. S. *et al.* The error distribution of BATSE gamma-ray burst locations. *Astrophys. J.* **122**, 503–518 (1999).
- Pendleton, G. N. *et al.* The structure and evolution of LOCBURST: The BATSE burst location algorithm. *Astrophys. J.* **512**, 362–376 (1999).
- Hakkila, J. *et al.* in *Gamma Ray Burst and Afterglow Astronomy 2001* (eds Ricker, G. R. & Vanderspek, R. K.) 176–178 (AIP Astronomy and Astrophysics Vol. 662, Woods Hole, Massachusetts, 2003).
- Balazs, L. G., Meszaros, A. & Horvath, I. Anisotropy of the sky distribution of gamma-ray bursts. *Astron. Astrophys.* **339**, 1–6 (1998).
- Magliocchetti, M., Ghirlanda, G. & Celotti, A. Evidence for anisotropy in the distribution of short-lived gamma-ray bursts. *Mon. Not. R. Astron. Soc.* **343**, 255–258 (2003).
- Rosswog, S. & Ramirez-Ruiz, E. On the diversity of short gamma-ray bursts. *Mon. Not. R. Astron. Soc.* **343**, L36–L40 (2003).
- Kalogera, V. *et al.* The cosmic coalescence rates of neutron star binaries. *Astrophys. J.* **601**, L179–L182 (2004).
- Butler, N. *et al.* GRB050709: A possible short-hard GRB localized by HETE. *GRB Circ. Netw.* **3570** (2005).
- Krimm, H. *et al.* GRB050724: Refined analysis of the Swift-BAT possible short burst. *GRB Circ. Netw.* **3667** (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the use of the publicly available BATSE Current Burst database. We thank M. Briggs for advice on BATSE instrumental characteristics. N.R.T., A.J.L. and R.S.P. acknowledge support from UK PPARC, and R.C. acknowledges the support of a University of Hertfordshire studentship.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.R.T. (nrt@star.herts.ac.uk).

LETTERS

An origin for short γ -ray bursts unassociated with current star formation

S. D. Barthelmy¹, G. Chincarini^{2,3}, D. N. Burrows⁴, N. Gehrels¹, S. Covino², A. Moretti², P. Romano², P. T. O'Brien⁵, C. L. Sarazin⁶, C. Kouveliotou⁷, M. Goad⁵, S. Vaughan⁵, G. Tagliaferri², B. Zhang⁸, L. A. Antonelli⁹, S. Campana², J. R. Cummings^{1,10}, P. D'Avanzo^{2,11}, M. B. Davies¹², P. Giommi¹³, D. Grupe⁴, Y. Kaneko¹⁴, J. A. Kennea⁴, A. King⁵, S. Kobayashi⁴, A. Melandri⁹, P. Meszaros^{4,15}, J. A. Nousek⁴, S. Patel¹⁴, T. Sakamoto^{1,10} & R. A. M. J. Wijers¹⁶

Two short (<2 s) γ -ray bursts (GRBs) have recently been localized^{1–4} and fading afterglow counterparts detected^{2–4}. The combination of these two results left unclear the nature of the host galaxies of the bursts, because one was a star-forming dwarf, while the other was probably an elliptical galaxy. Here we report the X-ray localization of a short burst (GRB 050724) with unusual γ -ray and X-ray properties. The X-ray afterglow lies off the centre of an elliptical galaxy at a redshift of $z = 0.258$ (ref. 5), coincident with the position determined by ground-based optical and radio observations^{6–8}. The low level of star formation typical for elliptical galaxies makes it unlikely that the burst originated in a supernova explosion. A supernova origin was also ruled out for GRB 050709 (refs 3, 31), even though that burst took place in a galaxy with current star formation. The isotropic energy for the short bursts is 2–3 orders of magnitude lower than that for the long bursts. Our results therefore suggest that an alternative source of bursts—the coalescence of binary systems of neutron stars or a neutron star-black hole pair—are the progenitors of short bursts.

On 2005 July 24 at 12:34:09.32 UT (Universal Time), the BAT instrument on the *Swift* observatory⁹ triggered and located GRB 050724 (ref. 6). The lightcurve for this event (Fig. 1) is somewhat different from the typical short GRB population as originally defined by the Burst and Transient Source Experiment (BATSE)¹⁰. Similar to GRB 050709 (ref. 11), the event consists of two parts that are spectrally very different. The main portion has a duration of 3.0 ± 1.0 s and is dominated by a hard spike lasting only 0.25 s. Following the main peak at time T , there is a second smaller peak at $T + 1.1$ s plus long-lasting faint emission out to $T + 140$ s, which is all in the 15–25 keV band. The faint tail images to the same location on the sky as the initial hard spike and is therefore definitely part of the initial event. The fluence in the faint tail is only 10% of that in the main peak.

The hardness ratio (50–100 keV to 25–50 keV) of the first peak is 0.91 ± 0.12 , which is on the low end of the distribution for short GRBs as detected with the BATSE instrument. The hardness ratio is 0.55 ± 0.23 for the second peak at $T + 1.1$ s and 0.31 ± 0.10 for the long-lasting emission out to $T + 140$ s. To determine whether this GRB would indeed have been classified as short in the BATSE sample, we performed a series of 100 simulations each, using the BATSE

large-area detector response matrices for three different angles of incidence, for all main constituents of GRB 050724. We found that in its typical trigger energy range (50–300 keV), BATSE would have triggered on the main hard peak and would definitely have detected the second pulse at $T + 1.1$ s; however, the softer pulse starting at $T + 30$ s would have been detected at best at the 0.3σ level and, thus, it would have not been accounted as part of the GRB. As a final test we passed the entire BAT lightcurve collected between 50–350 keV through the BATSE T_{90} algorithm, obtaining $T_{90} = 1.31 \pm 0.53$ s, entirely consistent with the short GRB population (but does not rule out a long-burst origin). Therefore, the long-lasting emission of this burst is probably a newly detected component of short bursts detected by BAT with its 15 keV threshold and not seen by BATSE with its >25 keV threshold (although possibly hinted at in the sum of multiple short BATSE GRBs^{12,13}).

The Swift X-Ray Telescope (XRT) began taking data 74 s after the BAT trigger⁶, and located a bright, fading X-ray source, identified as the X-ray afterglow^{14,15}. Ground-based observatories detected radio⁷ and optical^{16,17} afterglows within the XRT error circle and we obtained a high-precision X-ray position⁸ with our Chandra X-ray Observatory target of opportunity observation. The positions determined by the various instruments are given in Table 1 and shown in Fig. 2. The radio, optical, and X-ray positions are all coincident.

The fading afterglow of GRB 050724 was located¹⁸ approximately $1''$ south of the centre of (but within) a bright elliptical galaxy at a redshift of $z = 0.258$ (ref. 5). The projected offset from the centre of the galaxy corresponds to ~ 4 kpc (see Fig. 2). The host galaxy has an apparent magnitude of $K = 15.3$ within a $3''$ radius aperture¹⁷, which corresponds to a luminosity of $2 \times 10^{11} L_{\odot} \approx 1.7 L^*$ (where $L_{\odot}$ and L^* are the luminosities of the Sun and a typical galaxy, respectively). The total luminosity is probably slightly larger. Galaxies this luminous are relatively uncommon, with a comoving number density¹⁹ of only $3 \times 10^{-4} \text{ Mpc}^{-3}$. The probability that a GRB would occur randomly within $1''$ of the centre of a galaxy this luminous at a redshift of $z < 0.3$ is only $\sim 10^{-5}$. The small probability of an accidental superposition and lack of any other potential GRB host galaxies within the GRB error circle make the identification of this elliptical as the host galaxy quite secure.

The host galaxy's red colour²⁰, spectrum and visual appearance are all consistent with a luminous elliptical galaxy and very similar to the

¹NASA/Goddard Space Flight Center Greenbelt, Maryland 20771, USA. ²INAF—Osservatorio Astronomico di Brera, Via Bianchi 46, I-23807 Merate, Italy. ³Università degli studi di Milano Bicocca, Piazza delle Scienze 3, I-20126 Milano, Italy. ⁴Department of Astronomy and Astrophysics, Penn State University, University Park, Pennsylvania 16802, USA. ⁵Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. ⁶Department of Astronomy, University of Virginia, Charlottesville, Virginia 22903-0818, USA. ⁷NASA/Marshall Space Flight Center, NSSTC, XD-12, 320 Sparkman Drive, Huntsville, Alabama 35805, USA. ⁸Department of Physics, University of Nevada, Las Vegas, Las Vegas, Nevada 89154-4002, USA. ⁹Osservatorio Astronomico di Roma, Via di Frascati, 33 00040 Monte Porzio Catone, Italy. ¹⁰National Research Council, 2101 Constitution Avenue NW, Washington, DC 20418, USA. ¹¹Università degli Studi dell'Insubria, Dipartimento di Fisica e Matematica, via Valleggio 11, 22100 Como, Italy. ¹²Lund Observatory, Box 43, SE-221 00 Lund, Sweden. ¹³ASI Science Data Center, Via Galileo Galilei, I-00044 Frascati, Italy. ¹⁴Universities Space Research Association, NSSTC, XD-12, 320 Sparkman Drive, Huntsville, Alabama 35805, USA. ¹⁵Department of Physics, Penn State University, University Park, Pennsylvania 16802, USA. ¹⁶Astronomical Institute 'Anton Pannekoek', University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands.

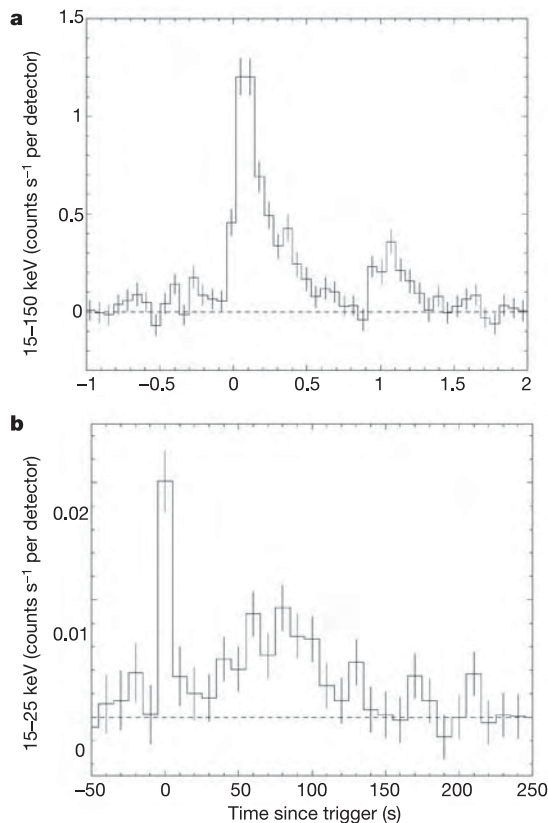


Figure 1 | BAT lightcurves for GRB 050724 showing the short duration of this GRB and the long softer emission. **a**, The prompt emission in the 15–150 keV energy band with a short-duration main spike of 0.25 s. T_{90} is 3.0 ± 1.0 s (T_{90} is the time during which 90% of the GRB photons are emitted¹⁰); the fluence is $(3.9 \pm 1.0) \times 10^{-7}$ erg cm⁻² and the peak flux is 3.5 ± 0.3 photons cm⁻² s⁻¹ (15–150 keV, 90% confidence level). **b**, Soft emission in the 15–25 keV energy band lasting >100 s (peak flux is $\sim 2 \times 10^{-9}$ erg cm⁻² s⁻¹). The error bars in both panels are one-sigma standard deviation. The BAT energy spectrum in the prompt portion ($T - 0.03$ to $T + 0.29$ s; where T equals BAT trigger time of 12:34:09.32 UT) is well fitted with a simple power-law model of photon index 1.38 ± 0.13 and normalization at 50 keV of 0.063 ± 0.005 photons cm⁻² s⁻¹ keV⁻¹ (15–150 keV, 90% confidence level). Count rate is normalized to a single detector of the 32,768 detectors in the full array of the BAT instrument.

properties of the host galaxy for the first localized short burst, GRB 050509B (ref. 1). One difference is that the host of GRB 050509B was located in a moderately rich cluster of galaxies, while the optical and X-ray observations of GRB 050724 suggest that this host elliptical is located in a lower-density region. The spectrum of the host shows no emission lines¹⁸ or evidence for recent star formation, and is consistent with a population of very old stars. This is true of most large elliptical galaxies in the present-day Universe, including the host galaxy of GRB 050509B. The elliptical hosts of these two short GRBs are very different from those for long bursts, which are typically sub-luminous, blue galaxies with strong star formation²¹.

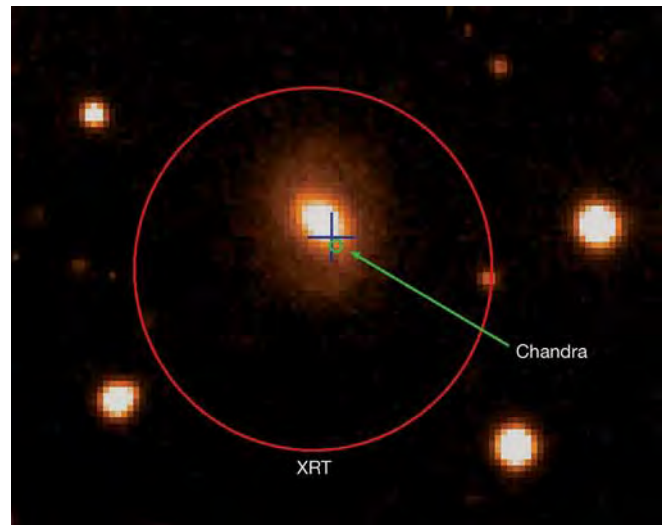


Figure 2 | VLT optical image¹⁷ showing the association of GRB 050724 with the galaxy. The blue cross is the position of the optical transient^{16,17}. The XRT (red circle) and Chandra (green circle) burst positions are superimposed on a bright red galaxy at redshift $z = 0.258$ (ref. 5), implying a low-redshift elliptical galaxy as the host. The XRT position has been further revised from the position of ref. 15 by astrometric comparison with objects in the field. The projected offset from the centre of the galaxy corresponds to ~ 4 kpc assuming the standard cosmology with $H_0 = 71$ km s⁻¹ Mpc⁻¹ and $(\Omega_M, \Omega_\Lambda) = (0.27, 0.73)$.

Thus the properties of these two short GRB hosts suggest that the parent populations and consequently the mechanisms for short and long GRBs are different in significant ways. Their non-star-forming elliptical hosts indicate that short GRBs could not have resulted from any mechanism involving massive star core collapse²² or recent star formation (for example, a young magnetar giant flare^{23,24}). As we previously noted¹, large elliptical galaxies are very advantageous sites for old, compact binary star systems, and thus good locations for neutron star–neutron star or neutron star–black hole mergers. Luminous elliptical galaxies are known to contain large populations of low-mass X-ray binaries containing neutron stars or black holes, and have large numbers of globular clusters within which compact binary stars can be formed dynamically with a much higher efficiency than in the field. Note, however, that mergers of compact objects are also expected to occur with a significant rate in star-forming galaxies; even if such mergers are the mechanism behind all short GRBs, one would not expect them all to occur in elliptical galaxies. In fact, the second short GRB with fine localization (GRB 050709)^{2–4} was in a star-forming galaxy at $z = 0.16$ and may be such a case.

Taking into account the host distance, we compare the energetics of short and long GRBs. The fluence in the first 3 s of emission is 6×10^{-7} erg cm⁻² in the 15–350 keV range, which translates roughly to a total 10 keV–1 MeV γ -ray fluence of $\sim 10^{-6}$ erg cm⁻². The fluences in the 30 to 200 s soft γ -ray peak and the X-ray afterglow are comparable at 7×10^{-7} erg cm⁻² and $\sim 10^{-6}$ erg cm⁻², respectively. These fluences are similar to those seen by BAT and other γ -ray detectors for long bursts. However, at a redshift of $z = 0.285$, the total

Table 1 | Position determinations for GRB 050724

Observatory	RA (J2000)	Dec. (J2000)	Error circle radius*	Notes	Ref.
Swift/BAT	16 h 24 min 43 s	−27° 31′ 30″	3′	1′ from Chandra position	6
Swift/XRT	16 h 24 min 44.41 s	−27° 32′ 28.4″	6″	Corrected astrometry relative to position in GCN Circular 3678	15
VLT	16 h 24 min 44.37 s	−27° 32′ 27″	0.5″		
VLA	16 h 24 min 44.37 s	−27° 32′ 27.5″	0.2″	One-sigma error	7
Chandra/ACIS	16 h 24 min 44.36 s	−27° 32′ 27.5″	0.5″		8

All the positions are consistent with each other to within the errors quoted for each. See Fig. 2. *90% confidence limit except for VLA. VLT, Very Large Telescope. VLA, Very Large Array. RA, right ascension; Dec., declination.

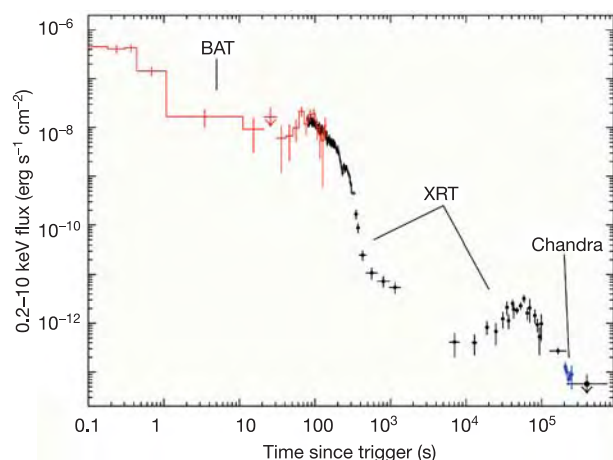


Figure 3 | The smooth transition of the GRB phase into the X-ray phase.

This shows the total X-ray afterglow lightcurve for GRB 050724 by combining data from BAT, XRT and Chandra. The prompt hard phase detected by BAT decays into the soft γ emission, and is followed by flaring ($\sim T + 80$ s) and then gradual fall-off into the X-ray afterglow measured by XRT. The BAT points are binned with variable time intervals such that the signal-to-noise ratio is >5 and $\Delta T < 10$ s. Using the Band model to simultaneously fit the BAT spectra in the 15–150 keV band in the $T + (50\text{--}150)$ s time interval with the XRT spectra in the 0.2–10 keV band in the $T + (79\text{--}150)$ s interval (yielding $E_{\text{break}} = 30$ keV, where E_{break} is the boundary energy between the two functional domains of the Band model), the BAT flux points (15–25 keV band) were scaled down into the X-ray band (unabsorbed 0.2–10 keV). The Band model was required because power-law fits to the BAT and XRT separately yielded Γ values of 2.5 and 1.7, respectively: that is, there was a spectral break between the two instruments. The XRT fluxes are unabsorbed values derived using best-fit XRT photon indices of $\Gamma = 1.9$ (for the ‘window timing’ mode data before 341 s) and $\Gamma = 1.5$ (for the ‘photon counting’ mode data after 342.9 s) and using the best-fit absorbing column density of $N_{\text{H}} = 5.9 \times 10^{21} \text{ cm}^{-2}$. All error bars are one-sigma standard deviation. The Chandra points are unabsorbed fluxes using the XRT-measured value for N_{H} and the Chandra-measured best-fit photon index of $\Gamma = 1.8$.

energy output of the prompt plus soft peak is $\sim 3 \times 10^{50}$ erg, which is significantly less than the $10^{52}\text{--}10^{54}$ erg isotropic-equivalent energy output of the long GRB class²⁵.

The early XRT lightcurve initially shows a steep decay with a slope of -2 . This component is connected to the soft γ -ray component around ~ 80 s identified in the extrapolated BAT lightcurve, which clearly overlaps and joins the early XRT measurements (Fig. 3). This flare-like event is followed by a very rapid decay after ~ 100 s (with index < -7), interrupted by a second, less-energetic flare around 200–300 s. The curve flattens at 700 s and has a roughly constant decay index all the way through the Chandra points to 2×10^5 s. A third significant flare starting at $\sim 2 \times 10^4$ s is superposed on this decay component, with the total energy even smaller (by a factor of ~ 3) than the second one. The steep decays following all three flares are too steep to be interpreted as the afterglow emission from a forward shock, but could be consistent with the high latitude emission ($\theta \gg 1/\Gamma$) from a fireball that suddenly stops radiating or goes off in an extremely low-density medium (naked GRBs)²⁶. This is pertinent for the late internal shock scenario as invoked to interpret the X-ray flares in long GRBs²⁷. This interpretation requires that the central engine remains active up to at least ~ 200 s (which is supported by the flaring event in the BAT-extrapolated X-ray lightcurve and the wiggles in the initial XRT lightcurve).

The current neutron star–neutron star merger models^{28,29} predict energy injection times much shorter than the >200 s time seen for GRB 050724. Black hole–neutron star mergers are more promising³⁰, because they allow for partial disruption of the neutron star and gradual accretion as the higher angular momentum material decays

through gravitational radiation, but even these models cannot extend the emission beyond a few tens of seconds. With the current Swift detection rate of one short burst per 2–3 months, the sample will quickly increase and will provide answers about how typical extended emission is for short burst.

Received 8 August; accepted 31 October 2005.

1. Gehrels, N. *et al.* A short γ -ray burst apparently associated with an elliptical galaxy at redshift $z = 0.225$. *Nature* **437**, 851–854 (2005).
2. Villaseñor, J. S. *et al.* Discovery of the short γ -ray burst GRB 050709. *Nature* **437**, 855–858 (2005).
3. Fox, D. B. *et al.* The afterglow of GRB 050709 and the nature of short-hard γ -ray bursts. *Nature* **437**, 845–850 (2005).
4. Covino, S. *et al.* Optical emission from GRB 050709: a short/hard GRB in a star-forming galaxy. *Astron. Astrophys.* (submitted).
5. Prochaska, J. X. *et al.* GRB 050724: Secure host redshift from Keck. *GRB Circ. Netw.* **3700** (2005).
6. Krimm, H. *et al.* GRB050724: Refined analysis of the Swift-BAT possible short burst. *GCN Circ.* **3667** (2005).
7. Berger, E. *et al.* The afterglow and elliptical host galaxy of the short γ -ray burst GRB 050724. *Nature* doi:10.1038/nature04238 (this issue).
8. Burrows, D. N. *et al.* GRB 050724: Chandra observations of the X-ray afterglow. *GCN Circ.* **3697** (2005).
9. Gehrels, N. *et al.* The Swift γ ray burst mission. *Astrophys. J.* **611**, 1005–1020 (2004).
10. Kouveliotou, C. *et al.* Identification of two classes of γ -ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
11. Boer, M. *et al.* GRB 050709: refined spectral and temporal properties. *GCN Circ.* **3653** (2005).
12. Connaughton, V. BATSE observations of γ -ray burst tails. *Astrophys. J.* **567**, 1028–1036 (2002).
13. Lazzati, D., Ramirez-Ruiz, E. & Ghisellini, G. Possible detection of hard X-ray afterglows of short γ -ray bursts. *Astron. Astrophys.* **379**, L39–L43 (2001).
14. Covino, S. *et al.* GRB 050724: a short burst detected by Swift. *GCN Circ.* **3665** (2005).
15. Antonelli, L. A. *et al.* GRB 050724: Swift XRT refined position. *GCN Circ.* **3678** (2005).
16. Gal-Yam, A. *et al.* GRB 050724: Optical variability in a nearby galaxy. *GCN Circ.* **3681** (2005).
17. D’Avanzo, P. *et al.* GRB050724: VLT observations of the variable source. *GCN Circ.* **3690** (2005).
18. Bloom, J. S. *et al.* GRB 050724: Optical imaging/astrometry. *GCN Circ.* **3672** (2005).
19. Kochanek, C. S. *et al.* The K-band galaxy luminosity function. *Astrophys. J.* **560**, 566–579 (2001).
20. Castro-Tirado, A. J. *et al.* GRB 050724: Optical and near-IR observations. *GCN Circ.* **3673** (2005).
21. Le Floc’h, E. *et al.* Are the hosts of γ -ray bursts sub-luminous and blue galaxies? *Astron. Astrophys.* **400**, 499–510 (2003).
22. Woosley, S. E. γ -ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).
23. Hurley, K. *et al.* An exceptionally bright flare from SGR 1806–20 and the origin of short duration γ -ray bursts. *Nature* **434**, 1098–1103 (2005).
24. Palmer, D. M. *et al.* A giant γ -ray flare from the magnetar SGR 1806–20. *Nature* **434**, 1107–1109 (2005).
25. Frail, D. A. *et al.* Beaming in γ -ray bursts: Evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
26. Kumar, P. & Panaitescu, A. Afterglow emission from naked γ -ray bursts. *Astrophys. J.* **541**, L51–L54 (2000).
27. Burrows, D. N. *et al.* Bright X-ray flares in γ -ray burst afterglows. *Science* **309**, 1833–1855 (2005).
28. Rosswog, S., Ramirez-Ruiz, E., & Davies, M. B. High-resolution calculations of merging neutron stars—III. γ -ray bursts. *Mon. Not. R. Astron. Soc.* **345**, 1077–1090 (2003).
29. Lee, W. H., Ramirez-Ruiz, E. & Granot, J. A compact binary merger model for the short, hard GRB 050509b. Preprint at (<http://arXiv.org/astro-ph/0504104>) (2005).
30. Davies, M. B., Levan, A. & King, A. The ultimate outcome of black hole–neutron star mergers. *Mon. Not. R. Astron. Soc.* **356**, 54–58 (2005).
31. Hjorth, J. *et al.* The optical afterglow of the short γ -ray burst GRB 050709. *Nature* **437**, 859–861 (2005).

Acknowledgements We acknowledge support from ASI, NASA and PPARC, and acknowledge benefits from collaboration within the EU FP5 Research Training Network ‘ γ -Ray Bursts: An Enigma and a Tool’.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.D.B. (scott@lheamail.gsfc.nasa.gov).

Chaos and threshold for irreversibility in sheared suspensions

D. J. Pine^{1,2}, J. P. Gollub^{1,3}, J. F. Brady⁴ & A. M. Leshansky⁵

Systems governed by time reversible equations of motion often give rise to irreversible behaviour^{1–3}. The transition from reversible to irreversible behaviour is fundamental to statistical physics, but has not been observed experimentally in many-body systems. The flow of a newtonian fluid at low Reynolds number can be reversible: for example, if the fluid between concentric cylinders is sheared by boundary motion that is subsequently reversed, then all fluid elements return to their starting positions⁴. Similarly, slowly sheared suspensions of solid particles, which occur widely in nature and science⁵, are governed by time reversible equations of motion. Here we report an experiment showing precisely how time reversibility⁶ fails for slowly sheared suspensions. We find that there is a concentration dependent threshold for the deformation or strain beyond which particles do not return to their starting configurations after one or more cycles. Instead, their displacements follow the statistics of an anisotropic random walk⁷. By comparing the experimental results with numerical simulations, we demonstrate that the threshold strain is associated with a pronounced growth in the Lyapunov exponent (a measure of the strength of chaotic particle interactions). The comparison illuminates the connections between chaos, reversibility and predictability.

The time reversibility of simple shear flows at low Reynolds number (Re) is demonstrated in a film by Taylor⁸. A drop of dye is placed in a fluid in the gap between concentric cylinders (a configuration known as circular Couette flow) and the inner cylinder rotated many turns. The drop seems to disappear as it spreads into a thin filament, but is reconstituted almost without change (except for slight blurring due to brownian motion) when the rotation is reversed. This behaviour is consistent with the time reversibility of the dynamical equations—the Stokes or ‘creeping flow’ equations⁴—that govern low Re flow. The same is not true for three-dimensional steady flows, which can exhibit chaotic advection and irreversibility⁹.

The reversible creeping flow equations also govern the low Re behaviour of a suspension of non-brownian particles (that is, particles too large to exhibit significant thermal motion). However, it was recently shown that in the creeping flow limit, the particle trajectories are chaotic¹⁰: that is, the suspended particles exhibit sensitive dependence on their initial (or current) positions. When the particle motion is described in a configuration space spanned by the $3N$ particle coordinates, nearby trajectories (in phase space) separate exponentially in time at a rate given by a positive Lyapunov exponent, λ . An earlier numerical study of only three particles falling through a viscous fluid also reveals chaotic particle dynamics¹¹. Thus, chaos is an intrinsic property of low Re flow of non-brownian suspensions.

In chaotic dynamical systems, trajectories are extremely sensitive to small perturbations from the deterministic paths dictated by

the dynamical equations. Since small perturbations are inevitably present, chaos may make it impossible for the system to retrace its microscopic dynamical path when reversed. Equivalently, chaos plus perturbations (noise) may lead to a loss of predictability after a finite time horizon¹². Normally, these properties are inferred from numerical solutions, since it is generally not possible to reconstitute the initial conditions in a real experiment, slightly perturb the system, and then follow its subsequent evolution. For a low Re non-brownian suspension, however, reversing the boundary motion is equivalent to reversing time. Thus, we have a unique opportunity to test the limits of reversibility and predictability experimentally: we can undertake a quantitative version of Taylor’s experiment and measure the degree to which particles return to their initial positions.

Previous experimental^{13–17} and numerical^{18,19} investigations show that the trajectories of non-brownian particles exhibit irregular and apparently random displacements, or effective diffusion, when a suspension is sheared unidirectionally between concentric cylinders with the inner one rotating (circular Couette flow). The origin of this irregular motion lies in the interactions between particles mediated by the fluid. Such interactions have been shown to produce a variety of important effects in suspensions, strongly affecting sedimentation^{20,21}, and causing particles to cluster²². However, the question of whether or not particle trajectories are reversible in one-dimensional shear flows has not been addressed experimentally.

Here we report a study of reversibility, which is made by straining a viscous suspension periodically in circular Couette flow. The time dependence of the strain γ , defined as the ratio of the azimuthal displacement of the inner cylinder to the width of the gap between the cylinders, is given by $\gamma(t) = \gamma_0 \sin \omega t$, where ω is the angular frequency of flow reversals and t is the time. The strain amplitude γ_0 is typically in the range 0.5–12. Straining the system to strain amplitude γ_0 evolves the system forward in time; by reversing the flow in the next quarter cycle, we can check to see if the particles return to their initial positions. Repeated sampling over many periods constitutes a severe test of reversibility, and allows the behaviour to be characterized statistically.

The suspension is composed of spherical polymethylmethacrylate (PMMA) particles of diameter $d = 230 \pm 20 \mu\text{m}$ dispersed in a multi-component fluid at a volume fraction ϕ in the range 0.1–0.4. The concentrations of the fluid components are adjusted²³ to match the density of the spheres (to avoid settling) and refractive index (to make the suspension transparent). A small number of particles are dyed black so that their positions can be tracked. The period of rotation is 5–100 s, and Re, based on the gap and the maximum fluid velocity, is held fixed at 0.001. Therefore, the low Re description applies.

Once oscillatory shear flow is established, the suspension is illuminated and the positions of the dyed particles recorded once

¹Department of Chemical Engineering and KITP, University of California, Santa Barbara, California 93106-5080, USA. ²Department of Physics and Center for Soft Matter Research, New York University, 2-4 Washington Place, New York, New York 10003, USA. ³Department of Physics, Haverford College, Haverford, Pennsylvania 19041, USA.

⁴Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA. ⁵Department of Chemical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel.

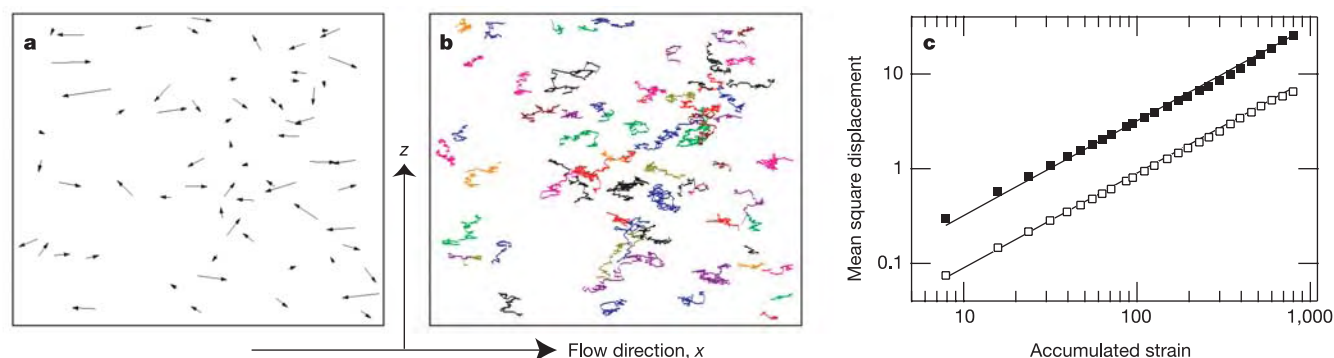


Figure 1 | Particle displacements and trajectories. **a**, Particle displacements in the x - z plane after one full cycle in a sheared suspension above the onset of irreversibility, amplified by a factor of 6 for clarity (volume fraction $\phi = 0.30$, strain amplitude $\gamma_0 = 2$). **b**, Some of the chaotic particle trajectories. **c**, Mean square particle displacements $\langle \Delta x^2 \rangle$ and $\langle \Delta z^2 \rangle$ after n full cycles as a function of the accumulated strain $\gamma = 4\gamma_0 n$ for $\phi = 0.40$ and $\gamma = 2.0$. The filled and open squares are the mean square displacements $\langle x^2 \rangle$ and $\langle z^2 \rangle$, respectively, obtained by averaging over particle trajectories such as those displayed in **b**; the solid lines through the data are least squares fits from which the diffusivities are determined. The fluctuations are anisotropic, growing more quickly along the flow direction (x) than along the axial direction (z). Experimental details: the diameter of the inner

cylinder of the Couette cell is 50 mm and the gap between the (concentric) cylinders is 2.5 mm; thus, a strain of 1 corresponds to an angular displacement of the inner cylinder of 5.7° . The PMMA particles have surface irregularities of only 2 nm, as measured by AFM. The fluid viscosity is 3 Pa s, about 3,000 times that of water. The suspension floats on a layer of mercury to eliminate end effects. The fractional accuracy of the phase at which the camera and frame grabber are triggered is typically better than 0.001, but the final results are not very sensitive to this quantity. We sample particle positions near the instant of maximum particle velocity. The particle displacements in the x and z directions after each full cycle are denoted by Δx and Δz , respectively.

per cycle using a camera and frame grabber. The total accumulated strain experienced by the particles over a run (taken to be positive at each instant) is $\gamma \equiv 4\gamma_0 n$, where n is the number of cycles, because the strain is γ_0 in each quarter cycle. To study the particle motion, we use tracking software²⁴ to locate the dark particles, whose coordinates x (along the direction of rotation) and z (along the cylinder axis) are tracked over the duration of the run. Sometimes particles are lost from view, but we typically track 60 particles or more over a total strain of 1,000 or so. This is ample to produce statistically accurate results.

We begin by describing the behaviour of the 30% concentration sample ($\phi = 0.30$). When γ_0 is less than about 1, all particles return to their original positions after each full cycle. On the other hand, if $\gamma_0 = 2$, the tracer particles fail to return by irregular amounts of typical magnitude $0.6d$ each cycle. These displacements are shown in Fig. 1a, and are larger along the flow (x) direction than along the axial (z) direction. The irreversibility after a full cycle, while comparable to the particle diameter, is substantially less than the typical displacement within a cycle. Trajectories sampled over many complete periods are shown in Fig. 1b; movies of both continuously and periodically sampled motion may be found in the Supplementary Information.

A statistical analysis of the motion, shown in Fig. 1c for $\phi = 0.40$, shows that the mean square particle displacements $\langle \Delta x^2 \rangle$ and $\langle \Delta z^2 \rangle$ after an integral number of cycles, averaged over many particles, are linear in the accumulated strain. The mean square displacements are also linear in time for a given driving frequency and strain amplitude, but what counts is the total strain. Data are essentially identical if Re is increased by a factor of 5; this shows that residual effects of the fluid inertia are not important. The anisotropy of shear flow leads to different magnitudes for $\langle \Delta x^2 \rangle$ and $\langle \Delta z^2 \rangle$. Anisotropy in particle transport also occurs for steady shearing, and is a consequence of ‘Taylor dispersion’²⁵.

We use these graphs of the mean square displacements to define non-dimensional effective diffusivities D_x and D_z along x and z , respectively, in the conventional way, but with total accumulated strain replacing time:

$$\langle (\Delta x/d)^2 \rangle = 2D_x\gamma \quad (1)$$

Equation (1) defines a dimensionless diffusivity that is independent

of the oscillation frequency, as we check empirically, thus allowing diffusivities measured at different frequencies and strain amplitudes to be compared meaningfully. Since the only natural time and length scales (other than the container dimensions) are the r.m.s. strain rate $\dot{\gamma}_{\text{rms}}$ and the particle diameter d , the dimensional particle diffusivity

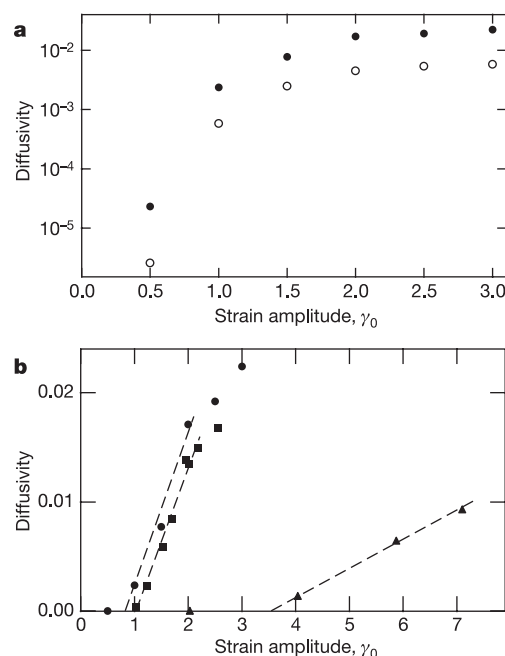


Figure 2 | Experimental diffusivities. **a**, Experimental effective diffusivities D_x (filled circles) and D_z (open circles) as a function of the oscillatory strain amplitude, on a logarithmic scale for volume fraction $\phi = 0.40$. The x and z diffusivities both become negligible for $\gamma_0 < 1$. **b**, Diffusivities D_x on a linear scale for three samples at different particle concentrations with linear extrapolations to zero diffusivity: triangles, $\phi = 0.20$; squares, $\phi = 0.30$; circles, $\phi = 0.40$. Data for D_z are omitted for clarity but extrapolate to the same values of the strain amplitude thresholds for the different concentrations.

D'_x defined by the dimensional form of equation (1), $\langle \Delta x^2 \rangle = 2D'_x t$, must scale like $D'_x \sim d^2 \dot{\gamma}_{\text{rms}}$.

The variation of the dimensionless diffusivities with strain amplitude for the 40% sample is shown in Fig. 2a. The measurement precision is about $\pm 10\%$. Both diffusivities become negligible (below our noise level) for $\gamma_0 < 0.8$. They appear to saturate for $\gamma_0 > 2$, though we cannot go higher than 3, where the large displacements make tracking difficult.

Irreversibility occurs only above a well-defined threshold strain amplitude, which depends strongly on the concentration of the suspension. This is shown in Fig. 2b, where we plot the dimensionless reversing diffusivities for three samples at different concentrations as a function of γ_0 . For $\phi = 0.40$, the threshold is close to $\gamma_0 = 0.8$, as can be seen from the linear extrapolation drawn through the circles in Fig. 2b, while for $\phi = 0.20$, the diffusivity becomes measurable only above $\gamma_0 = 3.5$. The strong concentration dependence of the extrapolated threshold is shown in Fig. 3.

These experiments raise a number of interesting questions. For example, what is the role of small imperfections and noise? Brownian motion would contribute displacements of only 10 nm in one period; the particle roughness of 2 nm is much smaller than the particle diameter (220,000 nm). Inertial effects have been shown experimentally to be negligible. Geometrical imperfections could play a role, but cannot explain a threshold.

Assuming that the particle interactions are chaotic and require interactions between at least three particles, as shown by János *et al.*¹¹, a strong concentration dependence of the threshold strain amplitude can be expected, since the probability of multiparticle encounters, and their strength, both depend strongly on concentration. Figure 3 shows that the threshold is roughly proportional to ϕ^{-2} experimentally and that numerical simulations (described later) show similar behaviour, though the data are limited.

Some insight can be obtained by comparing the experimental results to numerical simulations. We adapt the method of stokesian dynamics¹⁹ to simulate the reversing problem. Stokesian dynamics (SD) is akin to molecular dynamics except that the interactions are governed by the low Re or creeping flow equations. For sheared suspensions of neutrally buoyant particles, SD solves the Stokes equations for the fluid subject to the no slip boundary conditions on the surfaces of force-free and torque-free particles, to determine the particle velocities. The many-particle configuration $\mathbf{x}(t)$ evolves in time (neglecting inertia at small Re) according to

$$\frac{d\mathbf{x}}{dt} = \dot{\gamma}(t)\mathbf{U}(\mathbf{x}(t))d \quad (2)$$

where $\dot{\gamma}(t)$ is the instantaneous spatially constant shear rate and $\mathbf{U}(\mathbf{x}(t))$ are the configuration-dependent non-dimensional velocities of the particles. The SD method determines $\mathbf{U}$ for a given distribution

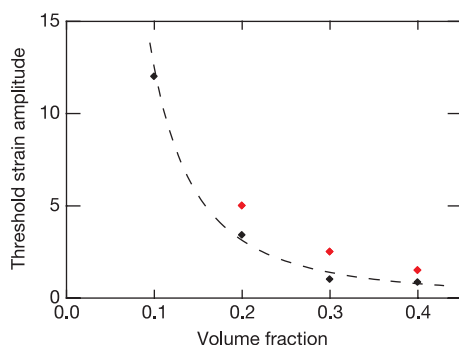


Figure 3 | Threshold strain amplitudes for the onset of irreversibility as a function of volume fraction. Experimental data, black diamonds; numerical simulation, red diamonds; power law fit to data $C\phi^{-\alpha}$, dashed line ($C = 0.14 \pm 0.03$, $\alpha = -1.93 \pm 0.14$). The threshold strain amplitudes are determined from the linear extrapolations in Fig. 2b.

of particles, and then uses equation (2) to evolve the particles in time. At each time step $\mathbf{U}$ is computed anew as the particle configuration changes.

Simulations are performed for $N = 512$ identical spherical particles subject to a bulk shear flow $\mathbf{u} = \dot{\gamma}(t)y\hat{\mathbf{e}}_x$, where $\hat{\mathbf{e}}_x$ is the unit vector in the flow direction and y the spatial coordinate in the velocity gradient direction. Periodic boundary conditions are used in all three directions to minimize boundary effects¹⁹. A square-wave protocol of period T is employed: for $0 < t < T/2$, $\dot{\gamma}(t) = \dot{\gamma}$ and for $T/2 < t < T$, $\dot{\gamma}(t) = -\dot{\gamma}$, which is then repeated. Simulations are performed for total accumulated strains $\gamma = \dot{\gamma}t$ between 400 and 1,000 and for different strain amplitudes $\gamma_0 = \dot{\gamma}T/4$. (The factor of 4 makes the experimental and numerical definitions correspond; the total strain per cycle is $4\gamma_0$ in both cases.)

The positions of all particles are followed in time; their displacements at the end of each period are recorded as a function of total accumulated strain. The mean square displacements exhibit linear growth with total strain at large total strain, just as in the experiments.

As shown in Fig. 4a, the simulations exhibit features similar to the experiments; the effective diffusivities are within a factor of 2–4 of the experiments, agreement that we consider reasonable given that the sources of irreversibility are different in the experiments and simulations. In particular, at small strain amplitudes ($\gamma_0 < 1$) the diffusivities from simulation are very small and of the order of the numerical accuracy of the calculations (which is limited by finite time steps, resolution, and so on). Therefore only qualitative agreement between simulation and experiment is expected.

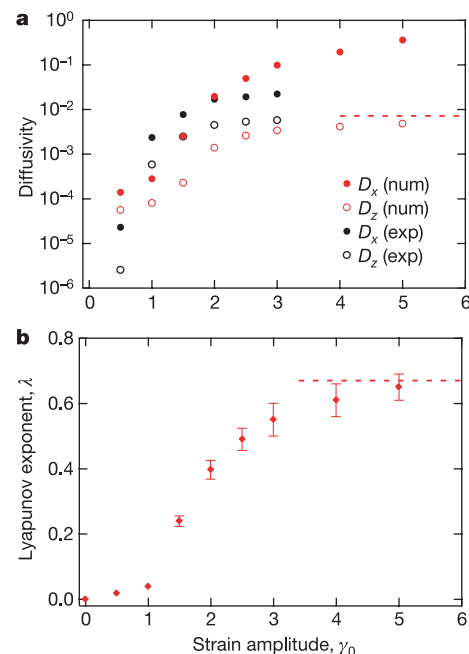


Figure 4 | Diffusivities and Lyapunov exponents. **a**, Numerically computed diffusivities (red symbols; 'num.') and experimental diffusivities (black symbols; 'exp.') for diffusion in the x (filled circles) and z (open circles) directions as a function of strain amplitude γ_0 for $\phi = 0.40$. In the simulations, the diffusivity D_y (not shown) in the y or velocity-gradient direction is also determined and is found to be intermediate between D_x and D_z . The reversing experiments for D_z at large strain amplitude compare well to steady shear simulations (dashed line), as is expected. **b**, Numerically computed Lyapunov exponent λ versus strain amplitude γ_0 . The rapid growth of λ with increasing γ_0 mirrors the rapid growth of the diffusivity in **a**, and shows that particle trajectories become increasingly sensitive to small perturbations as the strain amplitude increases. The values of λ for the reversing experiments approach the steady state value (dashed line) at large strain amplitude, as is expected. The error bars indicate the standard deviations found from fits to $\delta(t) \sim \exp(\lambda\dot{\gamma}t)$.

The most significant deviation between the simulations and the experiments is for the effective diffusivity at strain amplitudes beyond 2 (recall that the experiments are limited to strain amplitudes below 3 for the most concentrated samples). The experimental diffusivities appear to level off sooner for $\gamma_0 > 2$, which may be due to the finite gap size in the velocity gradient direction ($\sim 11d$) limiting particle excursions; in the simulations, the periodic boundary conditions allow the particles to wander off to infinity in all three directions.

Both the experiments and simulations exhibit a strain amplitude threshold below which the diffusivities drop dramatically. Experimentally, the critical strain amplitude rises from about 1 for $\phi = 0.3$ or 0.4 to more than 10 for $\phi = 0.1$, while in the simulations the threshold is somewhat higher, but still consistent with the ϕ^{-2} scaling mentioned earlier. Some insight concerning the origin of the rapid onset of diffusion with increasing strain amplitude can be gained by examining how rapidly small perturbations are amplified in an oscillating shear flow. To do this, we compute the evolution of two nearby $3N$ particle trajectories and determine their separation $\delta(t)$ in configuration space (spanned by all the $3N$ particle coordinates) following the method of Drazer *et al.*¹⁰. From a pair of SD runs, we find that nearby trajectories separate exponentially in time (or accumulated strain $\gamma = \dot{\gamma}t$) with a positive Lyapunov exponent λ defined by $\delta(t) \sim \exp(\lambda \dot{\gamma}t)$. These simulations reveal a rapid increase in λ with strain amplitude γ_0 , as shown in Fig. 4b, paralleling the behaviour of the diffusivities. The particle trajectories thus become increasingly sensitive to small perturbations as the strain amplitude increases; this supports the conclusion that the chaotic nature of the hydrodynamic interactions between particles leads to the observed irreversible behaviour.

The experiments show that the hydrodynamic interactions can lead to either reversible or irreversible motion, depending on the amount of deformation that is imposed on the suspension. Our numerical results demonstrate that the onset of irreversibility with increasing strain amplitude is associated with the rapid growth of the Lyapunov exponent, which measures the sensitivity of particle trajectories to small perturbations. It is striking that the onset is so marked and that it occurs at relatively small strain amplitudes in concentrated suspensions. Although the creeping flow equations are deterministic, the time (or strain) horizon of predictability is therefore quite limited, extending only out to a total strain of about unity (for concentrated samples). For larger strains, only a statistical description is possible, although important statistical quantities such as the diffusivities can be predicted from numerical solutions to the equations of motion.

These experiments and simulations provide a detailed picture of the onset of irreversibility, the loss of predictability, and the passage from a detailed deterministic description of a many-body dynamical system to a useful statistical description. Whereas some chaotic systems remain reversible for large relative displacements when accurately computed—for example, systems of planets or moons²⁶—suspensions become irreversible and unpredictable after quite small relative displacements. The large number of interacting particles may be responsible.

Received 8 August; accepted 25 October 2005.

1. Tolman, R. C. *The Principles of Statistical Mechanics* (Oxford Univ. Press, London, 1938).
2. Boffeta, G., Cencini, M., Falcioni, M. & Vulpiani, A. Predictability: a way to characterize complexity. *Phys. Rep.* **356**, 367–474 (2002).
3. Wang, G. M., Sevick, E. M., Mittag, E., Searles, D. J. & Evans, D. J. Experimental

demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Phys. Rev. Lett.* **89**, 050601 (2002).

4. Leal, L. G. *Laminar Flow and Convective Transport Processes: Scaling Principles and Asymptotic Analysis* (Butterworth-Heinemann, Boston, 1992).
5. Morrison, I. D. & Ross, S. *Colloidal Dispersions: Suspensions, Emulsions, and Foams* (Wiley-Interscience, Hoboken, 2003).
6. Lamb, J. S. W. & Roberts, J. A. G. Time-reversal symmetry in dynamical systems: a survey. *Physica D* **112**, 1–39 (1998).
7. Berg, H. C. *Random Walks in Biology* (Princeton Univ. Press, Princeton, New Jersey, 1993).
8. Taylor, G. I. & Friedman, J. *Low Reynolds Number Flows* (National Committee on Fluid Mechanics Films, Encyclopedia Britannica Educational Corp., United States, 1966); available as 'G. I. Taylor and Kinematic Reversibility' in Multi-Media Fluid Mechanics CD-ROM (eds Homsy, G. M.) (Cambridge Univ. Press, Cambridge, 2000).
9. Aref, H. Chaotic advection of fluid particles. *Phil. Trans. R. Soc. Lond. A* **333**, 273–288 (1990).
10. Drazer, D., Koplik, J., Khusid, B. & Acrivos, A. Deterministic and stochastic behaviour of non-Brownian spheres in sheared suspensions. *J. Fluid Mech.* **460**, 307–335 (2002).
11. János, I. M., Tél, T., Wolf, D. E. & Gallas, J. A. C. Chaotic particle dynamics in viscous flows: The three-particle Stokeslet problem. *Phys. Rev. E* **56**, 2858–2866 (1997).
12. Lighthill, J. The recently recognized failure of predictability in Newtonian dynamics. *Proc. R. Soc. Lond. A* **407**, 35–48 (1986).
13. Eckstein, E. C., Bailey, D. G. & Shapiro, A. H. Self-diffusion of particles in shear flow of a suspension. *J. Fluid Mech.* **79**, 191–208 (1977).
14. Leighton, D. & Acrivos, A. Measurement of shear-induced migration of particles in concentrated suspensions. *J. Fluid Mech.* **177**, 109–131 (1987).
15. Breedveld, V., van den Ende, D., Tripathi, A. & Acrivos, A. The measurement of the shear-induced particle and fluid tracer diffusivities by a novel method. *J. Fluid Mech.* **375**, 297–318 (1998).
16. Breedveld, V., van den Ende, D. & Jongschaap, R. Shear-induced diffusion and rheology of noncolloidal suspensions: Time scales and particle displacements. *J. Chem. Phys.* **114**, 5923–5936 (2001).
17. Zarraga, I. E. & Leighton, D. T. Measurement of an unexpectedly large shear-induced self-diffusivity in a dilute suspension of spheres. *Phys. Fluids* **14**, 2194–2201 (2002).
18. Marchioro, M. & Acrivos, A. Shear-induced particle diffusivities from numerical simulations. *J. Fluid Mech.* **443**, 101–128 (2001).
19. Sierou, A. & Brady, J. F. Shear-induced self-diffusion in non-colloidal suspensions. *J. Fluid Mech.* **506**, 285–314 (2004).
20. Segré, P. N., Herbolzheimer, E. & Chaikin, P. M. Long-range correlations in sedimentation. *Phys. Rev. Lett.* **79**, 2574–2577 (1997).
21. Segré, P. N., Liu, F., Umbanhowar, P. & Weitz, D. A. An effective gravitational temperature for sedimentation. *Nature* **409**, 594–597 (2001).
22. Voth, G. A. *et al.* Ordered clusters and dynamical states of particles in a vibrated fluid. *Phys. Rev. Lett.* **88**, 234301 (2002).
23. Krishnan, G. P., Beimef, S. & Leighton, D. T. Shear-induced radial segregation in bidisperse suspensions. *J. Fluid Mech.* **321**, 371–393 (1996).
24. Crocker, J. C. & Grier, D. G. Methods of digital video microscopy for colloidal studies. *J. Colloid Interf. Sci.* **179**, 298–310 (1996).
25. Taylor, G. I. Dispersion of soluble matter in solvent flowing slowly through a tube. *Proc. R. Soc. Lond. A* **219**, 186–203 (1953).
26. Sussman, G. J. & Wisdom, J. Chaotic evolution of the solar system. *Science* **257**, 56–62 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We appreciate discussions with L. G. Leal and G. Homsy. K. Knipmeyer and E. Knowlton provided assistance with data acquisition and reduction. This work was supported by the Keck Foundation (D.J.P.), the National Science Foundation (J.P.G.) and the US-Israel Binational Science Foundation (A.M.L.). The work was initiated during a granular physics workshop hosted by the Kavli Institute for Theoretical Physics at UCSB.

Author Contributions D.J.P. and J.P.G. were responsible for the experiments; J.F.B. and A.M.L. did the numerical simulations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.J.P. (pine@nyu.edu); correspondence concerning the simulations should be addressed to J.F.B. (jfb@cheme.caltech.edu).

Reorganization of the western Himalayan river system after five million years ago

Peter D. Clift¹ & Jerzy Blusztajn²

Uplift of mountains driven by tectonic forces can influence regional climate^{1,2} as well as regional drainage patterns, which in turn control the discharge of eroded sediment to the ocean^{3,4}. But the nature of the interactions between tectonic forces, climate and drainage evolution remains contested^{5–7}. Here we reconstruct the erosional discharge from the Indus river over the past 30 million years using seismic reflection data obtained from drill core samples from the Arabian Sea and neodymium isotope data. We find that the source of the Indus sediments was dominated by erosion within and north of the Indus suture zone until five million years ago; after that, the river began to receive more erosional products from Himalayan sources. We propose that this change in the erosional pattern is caused by a rerouting of the major rivers of the Punjab into the Indus, which flowed east into the Ganges river before that time. Seismic reflection profiles from the Indus fan suggest high mass accumulation rates during the Pleistocene epoch partly driven by increased drainage to the Indus river after five million years ago and partly by faster erosion linked to a stronger monsoon over the past four million years¹. Our isotope stratigraphy for the Indus fan provides strong evidence for a significant change in the geometry of western Himalayan river systems in the recent geologic past.

Tectonic forces in continental collision zones are widely recognized to drive rock and surface uplift that in turn affects climate locally, and even globally^{1,8}. Changing topography and climate also affects regional drainage patterns, which in turn control the discharge of eroded sediment to the ocean. Thus, mountain uplift and associated orographic rainfall might be expected to accelerate continental erosion and increase mass accumulation rates on continental margins. However, in practice the sediment discharge to the ocean may also be controlled by the loss or gain of drainage to the main river course. South and East Asia are the global type areas for studies of how tectonic forces have influenced climate and drainage evolution. In East Asia, Clark *et al.*³ analysed drainage patterns in eastern Tibet and suggested that the Red River had originally been the ancestral river of East Asia, suffering progressive loss of drainage to neighbouring systems, caused by the long-term change in regional topography. Unfortunately, the timing of capture events cannot be reconstructed from the terrestrial evidence alone.

Marine sediments can provide temporal control to drainage development models through the biostratigraphy that defines their depositional age. Capture events must affect the composition and rate of delivery of sediment to the ocean. In this study, we chose to study drainage evolution in the Indus river because the exhumation history of the source mountains is relatively well constrained⁹ and the palaeoceanography and palaeoclimatology of the Arabian Sea itself has been documented back to ~14 Myr ago, allowing ready comparison with the sediment record of erosion^{10,11}.

We reconstructed the long-term erosional discharge from the

Indus river using new and previously published seismic reflection data (Supplementary Fig. 1) to estimate accumulation rates of clastic material as a proxy for continental erosion rates. Although some sediment is preserved onshore, around two-thirds of the total volume is preserved offshore¹². Although the profiles used to generate this budget only extend ~400 km from the coast they provide representative coverage of the sub-seafloor structure over the thickest parts of the fan. They are dated by ties to industrial well Indus Marine A-1 (ref. 13). The budget calculation process is described in the Supplementary Methods. In addition, we studied sediment samples from a series of scientific and industrial boreholes across the Arabian Sea to characterize the changing composition of Indus discharge during mountain uplift (Fig. 1). Because no single borehole penetrates the entire sequence we used a series of wells to construct a composite section (Supplementary Table 1). The provenance of the sediment was constrained through Nd isotope analysis of bulk samples. This method is based on the age and compositional differences between rocks across the India–Asia collision zone^{14–17}.

The Nd isotope data show a coherent temporal development, with ϵ_{Nd} values¹⁸ rising gradually from around –11 at 30 Myr ago to –9.5 at 5 Myr ago (Fig. 2), implying a relatively stable provenance. This is surprising given the major tectonic events that occurred in the Himalayas during that interval, not least the exhumation of the Greater Himalayas, probably during the Early Miocene¹⁹. ϵ_{Nd} values of –10 to –11 are consistent with a relatively stable sediment provenance dominated by erosion from the Karakoram^{12,14}, and contrasting with the more negative (that is, Himalayan) values of the Bengal fan²⁰. Erosion from the Greater Himalayas is calculated to comprise ~25% of the total Indus discharge for ϵ_{Nd} values of –10. However, these values are less negative than those measured from sands in the modern delta (ϵ_{Nd} around –15)¹⁴. Our new results show that the low modern ϵ_{Nd} values represent a trend to more negative values that started after 5 Myr ago, reaching very low values no later than the Late Pleistocene (~300 kyr ago). This shift can only be achieved by greatly increasing the input from Himalayan sources at the relative expense of the Karakoram or arc rocks of the Indus suture zone. Although the difference between the modern river and the Pleistocene may partially reflect damming of the Indus river in the past 30 years, this first-order shift from the Oligo-Miocene values must be a response to natural processes.

The bulk of the Himalayan material now contributing to the Indus river is delivered by the four large rivers of the Punjab (the Sutlej, Ravi, Chennab and Jellum rivers; Fig. 1). Although erosion of the Nanga Parbat massif contributes material of very negative ϵ_{Nd} values (–22 to –30)²¹, the Indus river immediately downstream of Nanga Parbat has an ϵ_{Nd} value of around –11, which is far short of the –15 seen at the delta, or even in the Pleistocene fan sediments^{12,14}. The only way to shift the net Indus sediment budget from the pre-5-Myr levels to those seen today is by an increase in the relative discharge

¹Department of Geology and Petroleum Geology, University of Aberdeen, Aberdeen AB24 3UE, UK. ²Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

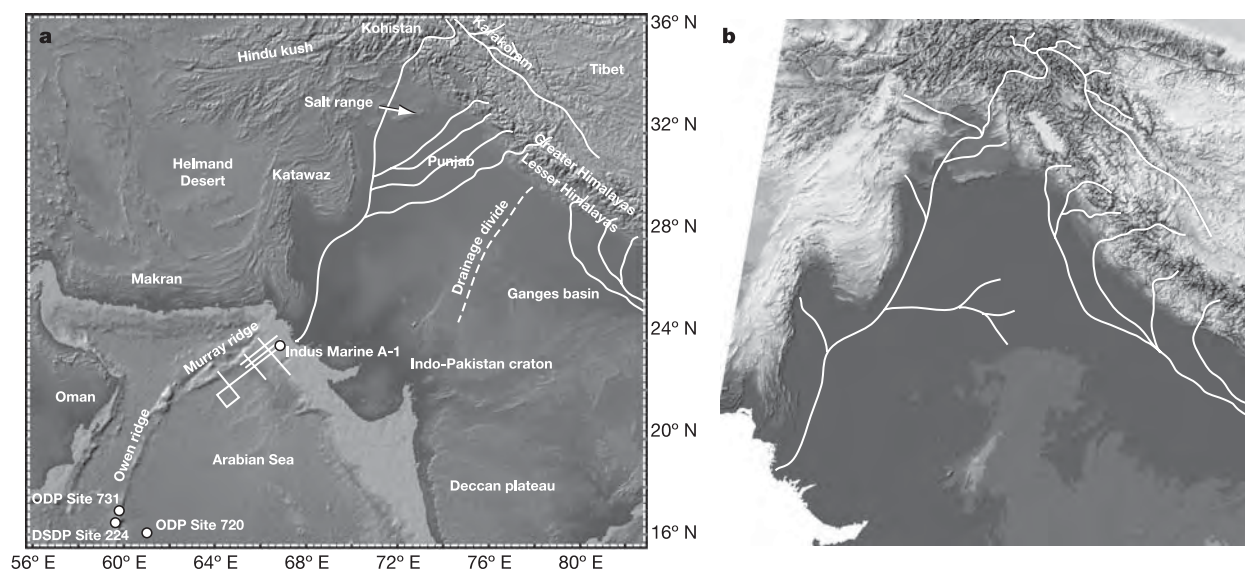


Figure 1 | Topographic maps showing post- and pre-capture Indus drainage, together with locations of drill and seismic data. a, Shaded relief map of the Arabian Sea and surrounding land masses, showing the location

of the drill sites and seismic profiles considered in this study. Major tributaries of the Indus river are shown. **b**, Proposed drainage geometry in the Indus river drainage before ~5 Myr ago.

from the Himalayas. This implies that input from the Punjabi rivers before 5 Myr must have been largely non-existent.

This seems odd, because reconstruction of the exhumation histories for the western Greater Himalayas shows that these mountains were in existence before 20 Myr ago^{9,21} and foreland sediments demonstrate that these ranges were being rapidly eroded during the Early to Mid-Miocene^{22,23}. Nonetheless, this erosion does not seem to be communicated to the Arabian Sea until after 5 Myr ago. We suggest that the simplest explanation for this pattern is that the ancestral Punjabi rivers were connected to the Ganges and not the Indus drainage. Support for this model comes from the palaeo-current measurements of Burbank *et al.*²³, who showed eastward flow in the foreland basin of northeast Pakistan, in the region now occupied by the Punjabi tributaries, during much of the Miocene.

The dramatic switch in isotopic character after 5 Myr ago must be driven by capture of these Punjabi tributaries to the Indus river. Why this capture occurred is not certain. This period was one of tectonic rejuvenation of the Main Central Thrust^{24,25}, which would be predicted to have increased the erosional contribution into the basin. In addition, early Pliocene uplift of the Salt range²⁶ may have been crucial in diverting rivers from their original southeastward flow into the Indus system. Alternatively, drainage capture may have occurred by a simple northeastward advance of tributaries of the Indus that short-circuited the older drainage pattern into the Ganges to provide a more direct route to the ocean.

The effect of the proposed capture on sediment accumulation rates can be seen in Fig. 2. Unfortunately the low resolution of the provenance and erosion histories makes a convincing link between the two impossible to establish at present. However, there is a clear increase in sedimentation rates from Pliocene to Pleistocene as the provenance was changing, consistent with drainage capture into the Indus system. Nd isotope character starts to change by 3.6 Myr ago when sedimentation rates were still low. However, the shift in isotope value is small, suggesting limited capture. The greatest change in provenance correlates at the first-order level with the increase in accumulation rates during the Pleistocene. However, increases in marine sedimentation rate are known from throughout Asia during the Pleistocene²⁷. Average sedimentation rates approximately doubled between the Pliocene and the Pleistocene in the Indus Fan, but how much of this increase reflects drainage capture is unclear.

The Nd isotope data allow us to generate a rough estimate of how much extra Himalayan discharge would be required to produce a shift in average ϵ_{Nd} values from -10 to -13 , assuming the discharge from the upper reaches of the Indus remained constant over the past 5 Myr. If the pre-capture Indus comprised ~25% Himalayan material (with an average ϵ_{Nd} value of -17 compared to -11 for the Karakoram and $+8$ for Kohistan¹⁴) to produce an average ϵ_{Nd} value of -10 , then the post-5-Myr-ago shift to values of -13 would require the proportion of Himalayan material to rise to ~70% of the total, approximately a threefold increase. Total sediment discharge would thus increase by around 45%. However, because mean

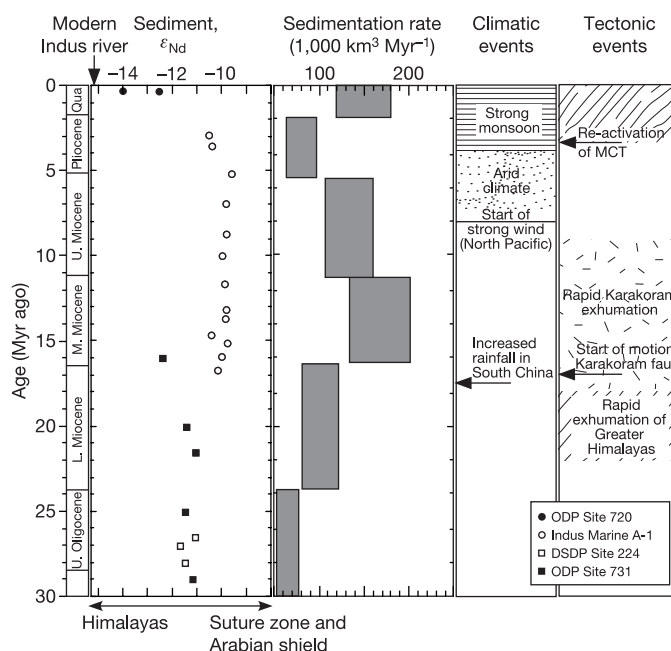


Figure 2 | The evolving Nd isotope composition and sedimentation rates on the Indus fan in relation to climatic and tectonic events known from onshore in Asia. MCT, Main Central Thrust. U., Upper; L., Lower; M., Middle; Qua, Quaternary.

accumulation rates approximately doubled after 5 Myr ago it follows that an additional volume of sediment, similar in size to that contributed from the Punjabi tributaries, must have been supplied to the Indus as a result of enhanced erosion throughout the catchment.

This study presents evidence for the nature and timing of a major drainage capture event in the western Himalayas. Faster sediment accumulation in the Pleistocene after a low in the Pliocene is estimated partly to reflect capture to the Indus after 5 Myr ago, but also seems to require faster erosion, probably driven by a stronger, wetter summer monsoon after 4 Myr ago¹. We note that fast erosion is not unique to the cyclical glacial–interglacial climate of the Plio–Pleistocene⁷, but must be controlled by other factors, at least during in the middle Miocene. Although the exact timing of drainage capture has yet to be established, the new isotope stratigraphy for the Indus fan provides strong evidence for a major change in the geometry of western Himalayan river systems in the relatively recent geologic past, probably caused by compressional deformation in the mountains. This study highlights the need to account for such capture events when trying to interpret marine erosional records, where changes in sediment composition and accumulation rates need not be directly linked to either source uplift or climate change, but simply to drainage reorganization.

Received 11 March; accepted 18 October 2005.

1. An, Z., Kutzbach, J. E., Prell, W. L. & Porter, S. C. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature* **411**, 62–66 (2001).
2. Prell, W. L. & Kutzbach, J. E. Sensitivity of the Indian monsoon to forcing parameters and implications for its evolution. *Nature* **360**, 647–652 (1992).
3. Clark, M. K. *et al.* Surface uplift, tectonics, and erosion of Eastern Tibet from large-scale drainage patterns. *Tectonics* **23**, TC1006 doi:10.1029/2002TC001402 (2004).
4. Brookfield, M. E. The evolution of the great river systems of southern Asia during the Cenozoic India–Asia collision; rivers draining southwards. *Geomorphology* **22**, 285–312 (1998).
5. Molnar, P. Nature, nurture and landscape. *Nature* **426**, 612–614 (2003).
6. Burbank, D. W. *et al.* Decoupling of erosion and precipitation in the Himalayas. *Nature* **426**, 652–655 (2003).
7. Zhang, P., Molnar, P. & Downs, W. R. Increased sedimentation rates and grain sizes 2–4 Myr ago due to the influence of climate change on erosion rates. *Nature* **410**, 891–897 (2001).
8. Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122 (1992).
9. Searle, M. P. in *Asian Tectonics* (eds An, Y. & Harrison, M.) 110–137 (Cambridge Univ. Press, Cambridge, UK, 1996).
10. Prell, W. L., Murray, D. W., Clemens, S. C. & Anderson, D. M. in *Synthesis of Results from Scientific Drilling in the Indian Ocean* (eds Duncan, R. A. *et al.*) 447–469 (American Geophysical Union Monograph 70, AGU, Washington DC, 1992).
11. Kroon, D., Steens, T. & Troelstra, S. R. Onset of monsoonal related upwelling in the western Arabian Sea as revealed by planktonic foraminifers. *Proc. ODP Sci. Res.* **117**, 257–263 (1991).
12. Clift, P. D. *et al.* Development of the Indus Fan and its significance for the erosional history of the western Himalaya and Karakoram. *Geol. Soc. Am. Bull.* **113**, 1039–1051 (2001).
13. Shuaib, S. M. Geology and hydrocarbon potential of offshore Indus Basin, Pakistan. *Am. Assoc. Petrol. Geol. Bull.* **66**, 940–946 (1982).
14. Clift, P. D. *et al.* Nd and Pb isotope variability in the Indus River system: Implications for sediment provenance and crustal heterogeneity in the Western Himalaya. *Earth Planet. Sci. Lett.* **200**, 91–106 (2002).
15. Debon, F., Le Fort, P., Sheppard, S. M. F. & Sonet, J. The four plutonic belts of the Transhimalaya–Himalaya; a chemical, mineralogical, isotopic, and chronological synthesis along a Tibet–Nepal section. *J. Petrol.* **27**, 219–250 (1986).
16. Petterson, M. G., Crawford, M. B. & Windley, B. F. Petrogenetic implications of neodymium isotope data from the Kohistan Batholith, North Pakistan. *J. Geol. Soc. Lond.* **150**, 125–129 (1993).
17. Whittington, A. *et al.* Lithostratigraphic correlations in the western Himalaya; an isotopic approach. *Geology* **27**, 585–588 (1999).
18. DePaolo, D. J. & Wasserburg, G. J. Nd isotopic variations and petrogenetic models. *Geophys. Res. Lett.* **3**, 249–252 (1976).
19. Hodges, K. V. Tectonics of the Himalaya and southern Tibet from two perspectives. *Geol. Soc. Am. Bull.* **112**, 324–350 (2000).
20. Bouquillon, A., France-Lanord, C., Michard, A. & Tiercelin, J. Sedimentology and isotopic chemistry of the Bengal Fan sediments: the denudation of the Himalaya. *Proc. ODP Sci. Res.* **116**, 43–58 (1990).
21. Walker, C. B., Searle, M. P. & Waters, D. J. An integrated tectonothermal model for the evolution of the High Himalaya in western Zaskar with constraints from thermobarometry and metamorphic modeling. *Tectonics* **20**, 810–833 (2001).
22. Najman, Y. *et al.* Early-middle Miocene paleodrainage and tectonics in the Pakistan Himalaya. *Geol. Soc. Am. Bull.* **115**, 1265–1277 (2003).
23. Burbank, D. W., Beck, R. A. & Mulder, T. in *Asian Tectonics* (eds An, Y. & Harrison, M.) 149–188 (Cambridge Univ. Press, Cambridge, UK, 1996).
24. Harrison, T. M. *et al.* A late Miocene–Pliocene origin for the central Himalayan inverted metamorphism. *Earth Planet. Sci. Lett.* **146**, 1–7 (1997).
25. Catlos, E. J. *et al.* Geochronologic and thermobarometric constraints on the evolution of the Main Central Thrust, central Nepal Himalaya. *J. Geophys. Res.* **106**, 16177–16204 (2001).
26. Burbank, D. W. & Beck, R. A. in *Tectonics of the Western Himalayas* (eds Malinconico, L. L. & Lillie, R. J.) 113–128 (Geological Society of America Special Paper 232, GSA, Boulder, Colorado, 1989).
27. Métivier, F., Gaudemer, Y., Tapponnier, P. & Klein, M. Mass accumulation rates in Asia during the Cenozoic. *Geophys. J. Int.* **137**, 280–318 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements P.D.C. thanks the Government of Pakistan and ConocoPhillips for access to the PAK77 survey data. We thank BGR (Hanover, Germany) for access to SO122 cruise data. N. Kusznir and A. Roberts are thanked for allowing us to use the Flex-Decomp backstripping software.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.D.C. (p.clift@abdn.ac.uk).

LETTERS

Efficacy of the post-perovskite phase as an explanation for lowermost-mantle seismic properties

James Wookey^{1†}, Stephen Stackhouse², J-Michael Kendall³, John Brodholt² & G. David Price²

Constraining the chemical, rheological and electromagnetic properties of the lowermost mantle (D'') is important to understand the formation and dynamics of the Earth's mantle and core. To explain the origin of the variety of characteristics of this layer observed with seismology, a number of theories have been proposed¹, including core–mantle interaction, the presence of remnants of subducted material and that D'' is the site of a mineral phase transformation. This final possibility has been rejuvenated by recent evidence for a phase change in MgSiO₃ perovskite (thought to be the most prevalent phase in the lower mantle²) at near core–mantle boundary temperature and pressure conditions³. Here we explore the efficacy of this 'post-perovskite' phase to explain the seismic properties of the lowermost mantle through coupled *ab initio* and seismic modelling of perovskite and post-perovskite polymorphs of MgSiO₃, performed at lowermost-mantle temperatures and pressures. We show that a post-perovskite model can explain the topography and location of the D'' discontinuity, apparent differences in compressional- and shear-wave models¹ and the observation of a deeper, weaker discontinuity^{4,5}. Furthermore, our calculations show that the regional variations in lower-mantle shear-wave anisotropy are consistent with the proposed phase change in MgSiO₃ perovskite.

The elastic and seismic properties of mineral crystals can be determined using *ab initio* methods, which solve (to a good approximation) Schrödinger's equation to model interactions in a system of nuclei and electrons. Of these, density functional theory⁶ (DFT) is one of the most reliable and efficient methods. DFT-based methods explicitly incorporating thermal effects have been applied to the perovskite and post-perovskite polymorphs of MgSiO₃ at lower-mantle pressures but primarily at low temperatures^{7–9}. Here we apply the technique of ref. 10 to MgSiO₃ at temperatures and pressures appropriate for the D'' region (see Fig. 1; see also the Supplementary Information for details of method, results and comparison with previous work^{8,9,11}). This provides us with the opportunity to assess directly the efficacy of a post-perovskite phase in explaining the seismic observations.

A sharp discontinuity in seismic velocity is observed in many places in the lowermost mantle (for a review see ref. 1). The depth of this feature varies considerably (150–450 km above the core–mantle boundary, CMB). On the basis of low-temperature analyses, it has been suggested that the transformation to a post-perovskite structure in MgSiO₃ is responsible for the observed seismic discontinuity^{7,9,12}, though this has been hitherto difficult to quantify and compare with the effects of pure thermal variation. Our *ab initio* simulations, however, give us the opportunity to do this: we can interpolate the five calculated sets of elastic parameters for each phase over a realistic

temperature and pressure regime for the lowermost mantle. We take three depth–temperature profiles from a mantle convection model⁴ (representing a hot, cold and average part of the lower mantle, see Fig. 1). We then convert the temperature profiles to velocity and density profiles, both with and without a transformation to post-perovskite (Fig. 2). The Clapeyron slope for the phase transformation has been calculated as 7.5 MPa K^{−1} (ref. 12), and 9.56 MPa K^{−1} (ref. 7). We use the latter value and a static transition pressure of 98.7 GPa (ref. 7).

The depth of the phase transition is strongly temperature-dependent, to the extent that it never occurs for the hot-region profile. This could explain the topography observed on the D'' discontinuity and the difficulty of confidently observing it in the regions of the lowermost mantle characterized by low seismic

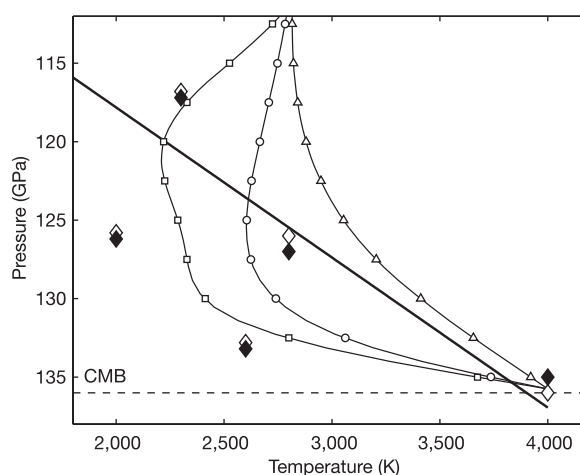


Figure 1 | Phase boundary⁷ and calculation points for the perovskite and post-perovskite polymorphs of MgSiO₃. We show the variation of temperature with pressure for three vertical profiles in a mantle convection model⁴. These come from 'hot' (triangles), 'cold' (squares) and 'average' (circles) parts of the mantle. Also plotted is the phase boundary from ref. 7; MgSiO₃ has a perovskite structure above this line, and a post-perovskite form below it. The 'cold' and 'average' profiles cross this boundary twice, thus predicting a thick post-perovskite layer underlain by a thin perovskite layer (similar to ref. 5). The 'hot' profile does not cross the phase boundary at all. It should be noted that this linear phase boundary is approximate; deviation from this will change the transition pressure. The dashed line indicates the CMB. The open and solid diamonds show the positions for which we have *ab initio* calculations of the seismic properties of the perovskite and post-perovskite, respectively.

¹School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK. ²Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK.

³Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK. [†]Present address: Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK.

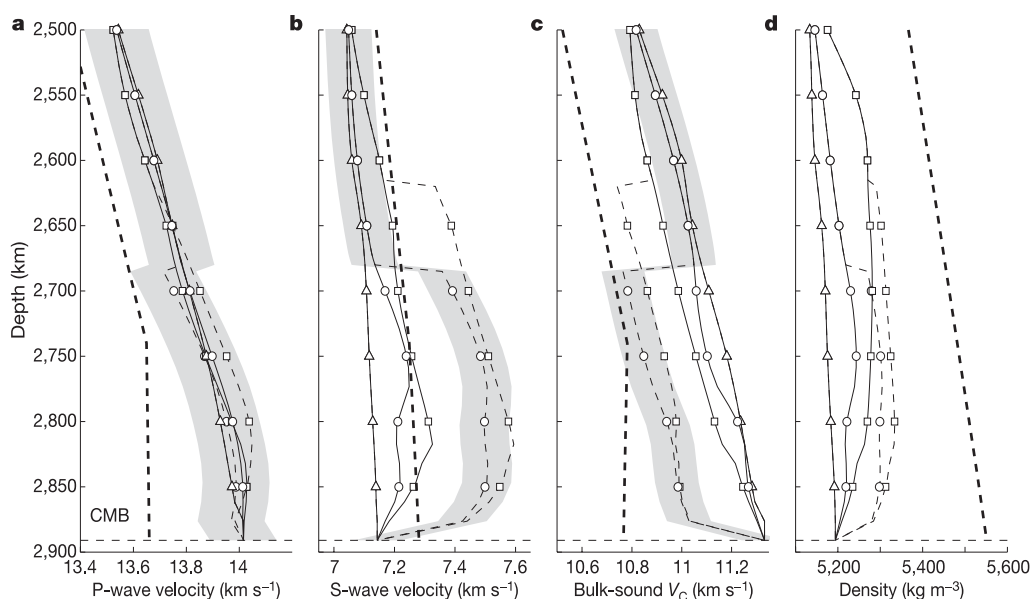


Figure 2 | Seismic properties interpolated from *ab initio* calculations. These show the P-wave (a) and S-wave (b) velocities, the bulk-sound speed (c) and the density (d) for the 'hot' (triangles), 'average' (circles) and 'cold' (squares) mantle convection profiles. These are converted from the temperature profiles in Fig. 1 into variation of the seismic properties with depth by interpolation of our *ab initio* results. The solid traces are the

variation solely due to temperature (in perovskite), and the dashed lines are where a phase-transformation is included. The shaded area shows the estimated constraint for the average profile (other profiles are similarly constrained). For comparison the velocities and density from the reference model ak135 (ref. 15) are included (black dashed trace).

velocities (such as the central Pacific)¹. The phase transition has different effects on the seismic velocities. The S-wave velocity (V_S) increases by around 4% at the discontinuity; this is enough to produce a strong shear-wave reflection. In contrast, the bulk-sound

speed (V_C) is as strongly reduced; this anticorrelation results in almost no P-wave velocity discontinuity. This is similar to previous athermal results^{7,9,11}. It is interesting to note that the strongest effect is in the 'average' profile, implying that it is not confined to the coldest

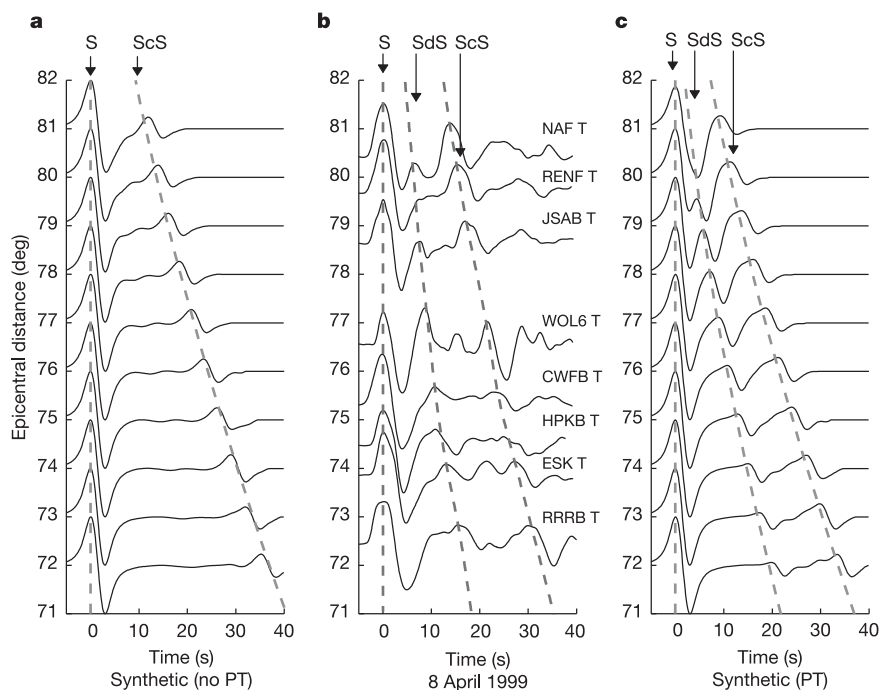


Figure 3 | Comparison between real and synthetic S-wave seismograms predicted from the velocity/density profiles shown in Fig. 2. b, Data from an earthquake on 8 April 1999 recorded at temporary stations (names as shown) in the SPICED array³⁰ and permanent stations in France and the UK. a, c, Synthetics generated using reflectivity modelling¹⁷: a shows the results of a purely thermal model; c includes the effects of a phase change. The upper and lower discontinuity depths are constrained to be those observed

for this event (approximately 250 km and 60 km above the CMB respectively⁴). Grey dashed lines show the travel times of the direct, D'' -reflected and core-reflected phases calculated using ray tracing. A D'' reflection is observed in the phase-change model that is very similar in character to the one observed for the data. This is absent from the thermal model. PT, phase transformation.

parts of the lowermost mantle. Furthermore, our results show that an anticorrelation of V_C and V_S in the lower mantle as seen in some tomography models^{13,14} might be explained by a phase change in MgSiO_3 . (The velocity anomalies in these models are relative to a global average at each depth, so warmer, perovskite-rich regions would appear fast in V_C and slow in V_S relative to the average, and vice versa for cooler, post-perovskite-rich regions).

The temperature–pressure curves for both ‘cold’ and ‘average’ mantle cross the phase boundary twice, providing a possible explanation for the second, deeper discontinuity which has been observed in the lowermost mantle^{4,5}. We predict this to occur at around 20 km above the CMB, whereas seismic observations suggest 60–80 km. This is, however, very sensitive to both the slope and intercept of the phase boundary and the temperature at the CMB (which is only known to within a few hundred Kelvin). We have also assumed a linear phase boundary, which may not be accurate. If the two observed discontinuities are related to this phase change, the topography on them should be anticorrelated; however, ascertaining whether this is the case requires more seismic observations.

The density profile shows a significant deviation from the average value given by the reference model ak135 (ref. 15). This is probably due to a combination of factors, but the most significant may be the absence of iron in the *ab initio* calculations. Incorporation of 15 mole per cent FeSiO_3 is sufficient to account for the density discrepancy, though also including less-dense¹⁶ MgO may increase this slightly ($\sim 1\text{--}2\%$). Differences between ak135 and the predicted seismic velocities are less than the estimated errors combined with the expected constraint in a one-dimensional reference model.

We apply reflectivity modelling¹⁷ to generate synthetic S-wave seismic sections (Fig. 3). With a phase-change model, an SdS reflection very similar to that observed in real data are evident. In P-wave synthetics no strong PdP phase is observed for either the thermal or phase-change model (see Supplementary Fig. S3). This is due to the anticorrelation of the bulk and shear moduli, which both contribute to P-wave velocity. However, the amplitude of this phase can be considerably enhanced by anisotropy in the lowermost mantle (see Supplementary Information and Supplementary Fig. S4), so observations of PdP could be consistent with a phase change model in regions of strong deformation. Our synthetics represent the seismic effect associated solely with the presence of a phase transformation, and they show that such a phenomenon effectively produces realistic seismic reflections.

The lowermost mantle also has a strong anisotropic signature. Seismic anisotropy (the variation of velocity with direction of wave propagation) provides a signature of deformation processes in the Earth. For example, measurements of anisotropy in the upper mantle have been used to study the dynamics of past and contemporary tectonics (see review in, for example, ref. 18). The anisotropy of the lowermost mantle has also been studied extensively (see reviews in refs 18 and 19). These studies have mainly restricted themselves to looking for radial anisotropy (also called vertical transverse isotropy, VTI), that is, measuring the difference in travel time between the horizontally (SH) and vertically polarized (SV) components of shear body-wave phases such as S, S_{diff} and ScS. In a recent study, a large data set of body-wave and surface-wave phases were inverted for shear-wave velocity and radial anisotropy, providing long-wavelength coverage of the whole mantle²⁰. This model shows a general trend of SH phases arriving before SV (that is, a horizontally polarized fast shear wave) in the faster regions of the lowermost mantle (which is inferred to be colder). This is also seen in globally averaged models²¹. This anisotropy is much smaller in the slower (and thus hotter) regions, such as the central Pacific, and does not show a consistent trend of fast-shear-wave orientation^{19,20}.

Several mechanisms have been suggested to explain observations of lowermost-mantle anisotropy. Perhaps the most obvious explanation is the lattice-preferred orientation (LPO) of the phases thought to comprise the bulk of the lower mantle: MgSiO_3 and

MgO (ref. 2). This requires a favourable deformation regime (dislocation creep as opposed to diffusion creep) to form an aggregate anisotropic medium. This has been observed experimentally at uppermost lower-mantle conditions²² for MgSiO_3 . However, the LPO of perovskite MgSiO_3 has not been thought to be a viable candidate mechanism, as it does not produce a horizontal fast-shear wave^{8,18}. The LPO of MgO (ref. 16) is more plausible, but it is a much less abundant phase. Another suggested mechanism is the shape-preferred orientation (SPO) of inclusions of a secondary phase. If this phase has strongly contrasting elastic properties to those of the matrix (for example, if the inclusion were remnant basaltic melt²³), then strong shear-wave anisotropy might result¹⁸.

The presence of post-perovskite MgSiO_3 provides a perhaps simpler explanation. The observed lateral variability of anisotropy in the lowermost mantle might be explained by thermal or compositional effects (or a combination of both). The *ab initio* calculations we have described provide the full elastic tensor for the perovskite and post-perovskite polymorphs of MgSiO_3 . These indicate that perovskite with a glide plane normal to the [001] axis²² shows a vertically polarized fast shear-wave. The slip systems of post-perovskite are unknown; however, with an glide plane normal to [001] post-perovskite shows a horizontally polarized fast shear-wave for all azimuths (Fig. 4), as is observed for the cold regions of the lowermost mantle²⁰. This analysis suggests that post-perovskite provides a good explanation for the anisotropy observed in the cold regions of the lowermost mantle. At the present time, it does not conflict with any of the observational or laboratory evidence. This does not imply that other mechanisms of anisotropy may not contribute, but they are not required by current observations.

The analyses we have presented show that a phase transition in MgSiO_3 is a compelling explanation for many of the seismic properties observed for the lowermost mantle. The sharp velocity change required to observe a reflected P and S phase is much easier to produce with a phase change than purely by thermal variation. The anisotropy observed in the cold part of the lowermost mantle can also be explained in this manner. The horizontal fast-shear wave we

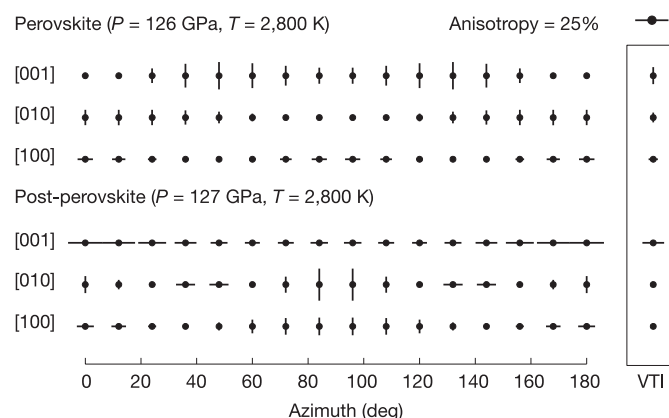


Figure 4 | Fast shear-wave orientation predicted for single-crystal perovskite and post-perovskite polymorphs of MgSiO_3 as a function of azimuth. The ray paths of the phases generally used to observe lowermost-mantle anisotropy¹⁹ (S, ScS, S_{diff}) are very close to horizontal at the base of the mantle. Each black vector indicates the orientation of the fast shear-wave for a given (horizontal) propagation direction, scaled by anisotropy magnitude. The lowermost mantle seems to be predominantly transversely isotropic, with a horizontally polarized fast shear wave^{19–21} (that is, $V_{\text{SH}} > V_{\text{SV}}$). For perovskite with a [001] glide plane (as is observed for the top of the lower mantle²²), a vertical fast shear-wave is predicted. A horizontal fast shear-wave is predicted for post-perovskite with a [001] glide plane. Perovskite with a [100] glide plane also shows $V_{\text{SH}} > V_{\text{SV}}$, but the anisotropy is much smaller. This implies that post-perovskite at realistic conditions is a much better candidate for explaining lowermost-mantle anisotropy.

predict for post-perovskite is in agreement with observations, and the absence of the phase in the hotter regions may explain why no consistent trend is seen beneath the central Pacific¹⁹. The apparent anticorrelation between bulk-sound speed and shear-wave velocity anomalies reported for the lower mantle^{13,14} can also be explained with a phase transformation, although this does not preclude a contribution of chemical origin. If present, chemical heterogeneity might also account for small-scale features such as basal ultralow-velocity zones²⁴ and scatterers^{25,26}.

We argue that post-perovskite MgSiO₃ is therefore likely to be both present and relatively abundant in the cold regions of the lowermost mantle, though we do not exclude the possibility of an additional source of chemical heterogeneity such as the presence of remnant slab material or core intrusion. The lowermost mantle is an important boundary to both mantle²⁷ and core convection²⁸, and the presence of this post-perovskite has implications for both. Convection in the core may be influenced by the mantle in several ways, and the thermal and electrical conductivity of D'' are an important and as-yet-unknown quantity. The presence of a dense basal layer may also significantly affect the style of convection in the mantle, controlling the timescale and morphology of features such as plumes²⁹.

Received 26 July; accepted 18 October 2005.

- Wysession, M. E., et al. in *The Core-Mantle Boundary Region* (eds Gurnis, M., Wysession, M. E., Knittle, E. & Buffett, B. A.) Vol. 28, 273–297 (American Geophysical Union, Washington DC, 1998).
- Ringwood, A. E. Phase transformations and their bearing on the constitution and dynamics of the mantle. *Geochim. Cosmochim. Acta* **55**, 2083–2110 (1991).
- Murakami, M., Hirose, K., Kawamura, K., Sata, N. & Ohishi, Y. Post-perovskite phase transition in MgSiO₃. *Science* **304**, 855–857 (2004).
- Thomas, C., Kendall, J.-M. & Lowman, J. Lower-mantle seismic discontinuities and the thermal morphology of subducted slabs. *Earth Planet. Sci. Lett.* **225**, 105–113 (2004).
- Herlund, J. W., Thomas, C. & Tackley, P. J. A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle. *Nature* **434**, 882–886 (2005).
- Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B* **136**, 864–871 (1964).
- Oganov, A. R. & Ono, S. Theoretical and experimental evidence for a post-perovskite phase of MgSiO₃ in Earth's D'' layer. *Nature* **430**, 445–448 (2004).
- Wentzcovitch, R. M., Karki, B. B., Karato, S. & Da Silva, C. R. S. High pressure elastic anisotropy of MgSiO₃ perovskite and geophysical implications. *Earth Planet. Sci. Lett.* **164**, 371–378 (1998).
- Tsuchiya, T., Tsuchiya, J., Umemoto, K. & Wentzcovitch, R. M. Elasticity of post-perovskite MgSiO₃. *Geophys. Res. Lett.* **31**, L14603 (2004).
- Stackhouse, S., Brodholt, J. P., Price, G. D., Wookey, J. & Kendall, J.-M. The effect of temperature on the acoustic anisotropy of the perovskite and post-perovskite polymorphs of MgSiO₃. *Earth Planet. Sci. Lett.* **230**, 1–10 (2005).
- Iitaka, T., Hirose, K., Kawamura, K. & Murakami, M. The elasticity of the MgSiO₃ post-perovskite phase in the Earth's lowermost mantle. *Nature* **430**, 442–445 (2004).
- Tsuchiya, T., Tsuchiya, J., Umemoto, K. & Wentzcovitch, R. M. Phase transition in MgSiO₃ perovskite in the Earth's lower mantle. *Earth Planet. Sci. Lett.* **224**, 241–248 (2004).
- Masters, G., Laske, G., Bolton, H. & Dziewonski, A. M. in *Earth's Deep Interior: Mineral physics and Tomography From the Atomic to the Global Scale* (eds Karato, S., Forte, A., Liebermann, R., Masters, G. & Stixrude, L.) Vol. 117, 201–213 (American Geophysical Union, Washington DC, 2000).
- Trampert, J., Deschamps, F., Resovsky, J. & Yuen, D. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle. *Science* **306**, 853–856 (2004).
- Kennett, B. L. N., Engdahl, E. R. & Buland, R. Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.* **122**, 108–124 (1995).
- Karki, B. B., Wentzcovitch, R. M., Gironcoli, S. d. & Baroni, S. First-principles determination of elastic anisotropy and wave velocities of MgO at lower mantle conditions. *Science* **286**, 1705–1707 (1999).
- Mueller, G. The reflectivity method; a tutorial. *Z. Geophys.* **58**, 153–174 (1985).
- Kendall, J.-M. in *Earth's Deep Interior: Mineral physics and Tomography From the Atomic to the Global Scale* (eds Karato, S., Forte, A., Liebermann, R., Masters, G. & Stixrude, L.) Vol. 117, 133–159 (American Geophysical Union, Washington DC, 2000).
- Lay, T., Williams, Q., Garnero, E. J., Kellogg, L. & Wysession, M. E. in *The Core-Mantle Boundary Region* (eds Gurnis, M., Wysession, M. E., Knittle, E. & Buffett, B. A.) Vol. 28, 299–318 (American Geophysical Union, Washington DC, 1998).
- Panning, M. & Romanowicz, B. Inferences on flow at the base of Earth's mantle based on seismic anisotropy. *Science* **303**, 351–353 (2004).
- Montagner, J.-P. & Kennett, B. L. N. How to reconcile body-wave and normal-mode reference Earth models. *Geophys. J. Int.* **125**, 229–248 (1996).
- Cordier, P., Ungar, T., Zsoldos, L. & Tichy, G. Dislocation creep in MgSiO₃. *Nature* **428**, 837–840 (2004).
- Hirose, K., Fei, Y., Ma, Y. & Mao, H.-K. The fate of subducted basaltic crust in the Earth's lower mantle. *Nature* **397**, 53–56 (1999).
- Garnero, E. J., Revenaugh, J., Lay, T. & Kellogg, L. H. in *The Core-Mantle Boundary Region* (eds Gurnis, M., Wysession, M. E., Knittle, E. & Buffett, B. A.) Vol. 28, 319–334 (American Geophysical Union, Washington DC, 1998).
- Thomas, C., Weber, M., Wicks, C. W. & Scherbaum, F. Small scatterers in the lower mantle observed at German broadband arrays. *J. Geophys. Res.* **104**, 15073–15088 (1999).
- Brana, L. & Helffrich, G. A scattering region near the core-mantle boundary under the North Atlantic. *Geophys. J. Int.* **158**, 625–636 (2004).
- Tackley, P. J. in *The Core-Mantle Boundary Region* (eds Gurnis, M., Wysession, M. E., Knittle, E. & Buffett, B. A.) Vol. 28, 231–253 (American Geophysical Union, Washington DC, 1998).
- Gubbins, D. Geomagnetic polarity reversals: a connection with secular variation and core-mantle interaction? *Rev. Geophys.* **32**, 61–83 (1994).
- Jellinek, A. M. & Manga, M. The influence of a chemical boundary layer on the fixity, spacing and lifetime of mantle plumes. *Nature* **418**, 760–763 (2002).
- Kendall, J.-M. & Helffrich, G. SPICED: imaging the deep Earth. *Astron. Geophys.* **42**, 26–29 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Dobson for discussions. This work was supported by the Deep Earth System NERC consortium grant.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.W. (j.wookey@bristol.ac.uk).

LETTERS

The earliest record of human activity in northern Europe

Simon A. Parfitt^{1,2}, René W. Barendregt³, Marzia Breda⁴, Ian Candy^{5,6}, Matthew J. Collins⁷, G. Russell Coope^{6,8}, Paul Durbidge⁹, Mike H. Field¹⁰, Jonathan R. Lee^{6,11}, Adrian M. Lister¹², Robert Mutch⁹, Kirsty E. H. Penkman⁷, Richard C. Preece¹³, James Rose⁶, Christopher B. Stringer^{2,6}, Robert Symmons², John E. Whittaker², John J. Wymer¹⁴ & Anthony J. Stuart^{12,15}

The colonization of Eurasia by early humans is a key event after their spread out of Africa, but the nature, timing and ecological context of the earliest human occupation of northwest Europe is uncertain and has been the subject of intense debate¹. The southern Caucasus was occupied about 1.8 million years (Myr) ago², whereas human remains from Atapuerca-TD6, Spain (more than 780 kyr ago)³ and Ceprano, Italy (about 800 kyr ago)⁴ show that early *Homo* had dispersed to the Mediterranean hinterland before the Brunhes–Matuyama magnetic polarity reversal (780 kyr ago). Until now, the earliest uncontested artefacts from northern Europe were much younger, suggesting that humans were unable to colonize northern latitudes until about 500 kyr ago^{5,6}. Here we report flint artefacts from the Cromer Forest-bed Formation at Pakefield (52° N), Suffolk, UK, from an interglacial sequence yielding a diverse range of plant and animal fossils. Event and lithostratigraphy, palaeomagnetism, amino acid geochronology and biostratigraphy indicate that the artefacts date to the early part of the Brunhes Chron (about 700 kyr ago) and thus represent the earliest unequivocal evidence for human presence north of the Alps.

The Cromer Forest-bed Formation (CF-bF), exposed discontinuously for a distance of more than 80 km along the North Sea coast of eastern England, has long been famous for its early Middle Pleistocene fossils^{7–11}. Spectacular finds include many extinct large mammals, molluscs, beetles, remains of fruits, seeds and even trees from which the deposits get their name. Recent work on vertebrate and molluscan faunas has shown that the CF-bF is much more complex than previously realized and includes evidence for at least six distinct temperate phases between about 780 and 450 kyr ago¹². The CF-bF sediments are primarily organic detritus muds and sands laid down within channels and on the floodplains of rivers, which drained central and eastern England before ice sheets invaded the area about 450 kyr ago¹³. During this time, Britain was connected to what is now mainland Europe, East Anglia being located at the southwestern margin of a large coastal embayment around the subsiding North Sea basin. The flint artefacts found at Pakefield (52° 25.9' N, 1° 43.8' W) were recovered from river sediments, with a significant quartz and quartzite component, which formed the floodplain of the lower reaches of the erstwhile Bytham River that drained the English Midlands at this time (Fig. 1)^{13,14}.

Despite two centuries of investigation, no convincing artefacts have hitherto been found stratified within the CF-bF. However, several excavations in the recently re-exposed coastal sections of CF-bF (and associated deposits) between Pakefield and Kessingland^{11,14–16} resulted in the discovery of 32 worked flints (length more than 20 mm), including a simple flaked core, a crudely retouched flake and a quantity of waste flakes (Fig. 2). These artefacts come from clear stratigraphical contexts and are associated with a wealth of evidence that allows a rare opportunity to reconstruct the environment inhabited by the humans who made the tools (see Supplementary Table 1). The artefacts, all in very sharp condition, are made of good quality black flint. Unworked surfaces, where present, are water-worn, indicating that the raw material might have been collected from the adjacent river channel. The assemblage lacks formal tools and is thus consistent with a Mode 1 technology (that is, flakes, pebble tools and choppers made with hard hammers¹⁷), but this interpretation must remain provisional in view of the small sample size.

The artefacts came from four different contexts within an interglacial infill of a channel incised into Early Pleistocene marine sediments and overlain by a sequence of marine sands, glaciofluvial sediments and Lowestoft Till (Fig. 3). The oldest artefact was found in the upper levels of the estuarine silts (Fig. 3) containing marine and brackish-water ostracods, foraminifera and sparse marine mammals (including dolphin and walrus). All other artefacts were found in the CF-bF, which comprises the 'Rootlet bed', 'Unio-bed' and 'laminated silts'^{14–16}. Two artefacts were found in overbank sediments with well-developed soil features including numerous fossil root-casts and pedogenic carbonate nodules ('Rootlet bed'). Most of the artefacts ($n = 28$) came from a lag gravel ('Unio-bed') at the base of the laminated silts that fill the channel cut into the overbank sediments, but a single flake (Fig. 2c) was found in the laminated silts at the edge of the channel. Sediments overlying the channel infill were originally interpreted as glaciofluvial deposits attributed to the North Sea Drift Formation (Marine Isotope Stage (MIS) 12; refs 11, 18; Fig. 3b), thus implying a long hiatus between the CF-bF and the glacial sequence. A new interpretation¹⁴ suggests that a more complete sequence is represented (Fig. 3b) by the complex succession of marine sands, Bytham River sands and gravels, and glaciofluvial sands and gravels from the Happisburgh Glaciation,

¹Institute of Archaeology, University College London, London WC1H 0PY, UK. ²Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. ³University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada. ⁴Dipartimento di Geologia, Paleontologia e Geofisica, University of Padova, Via Giotto 1, 35100 Padova, Italy. ⁵Department of Geography, Loughborough University, Loughborough LE11 3TU, UK. ⁶Department of Geography, Royal Holloway, University of London, Egham TW20 0EX, UK. ⁷BioArch, Biology S Block, University of York, PO Box 373, York YO10 5YW, UK. ⁸Department of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ⁹Lowestoft Museum, Oulton Broad, Lowestoft NR33 9JR, UK. ¹⁰Church Cottage, Church Street, Fenny Compton, Warwickshire CV47 2YE, UK. ¹¹British Geological Survey, Keyworth, Nottingham NG12 5GG, UK. ¹²Department of Biology, University College London, Gower Street, London WC1E 6BT, UK. ¹³Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. ¹⁴17 Duke Street, Bileston, Ipswich IP7 7EW, UK. ¹⁵University of Durham School of Biological and Biomedical Sciences, Durham DH1 3LE, UK.

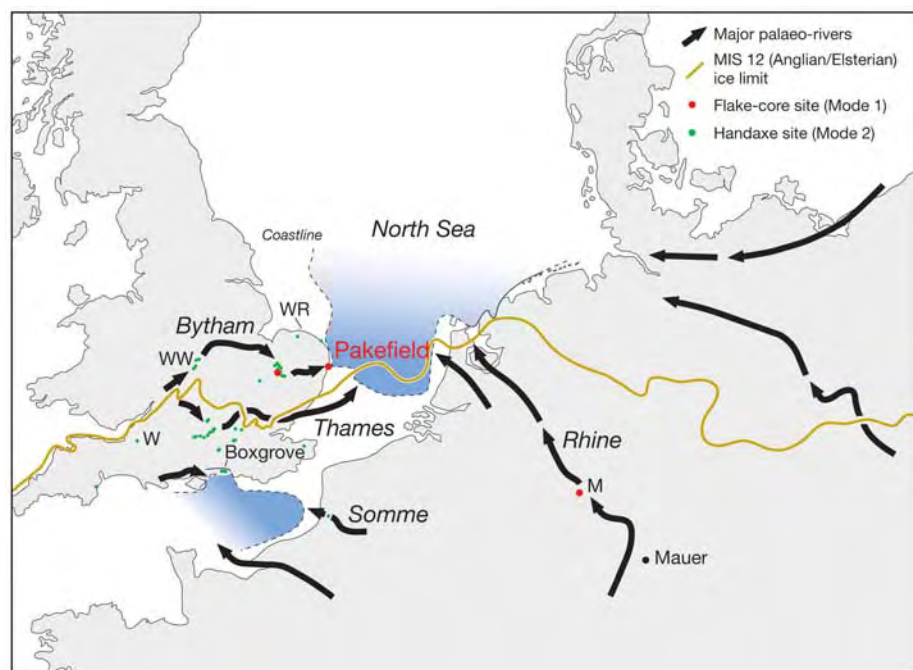


Figure 1 | A reconstruction of the palaeogeography of northwest Europe during the early Middle Pleistocene^{10,13,14,17,24,30}. All known Anglian/Elsterian and earlier archaeological sites are shown, and their concentration in southern England is highlighted. These sites can now be shown to provide a record of intermittent early human occupation of about 300 kyr duration. Human remains are known from Mauer and Boxgrove. M, Miesenheim I; W, Westbury-sub-Mendip; WR, West Runton; WW, Waverley Wood. For clarity, the names of rivers or geographical features are italicized and site names are in Roman font.

which are separated from the CF-bF by a periglacial interval indicated by the presence of frost-cracks. Another marine deposit overlies these sands and gravels and the whole sequence is capped by the Lowestoft Till Member, deposited by the Anglian Glaciation¹¹, when ice sheets reached their maximum extent in Britain.

Fossils from sediments that contain the artefacts indicate that the local climate was significantly different from that of the present day in terms of temperature and seasonality of precipitation. Plant (*Trapa natans*, *Salvinia natans* and *Corema album*) and beetle (*Cybister lateralis*, *Oxytelus opacus* and *Valgus hemipterus*)

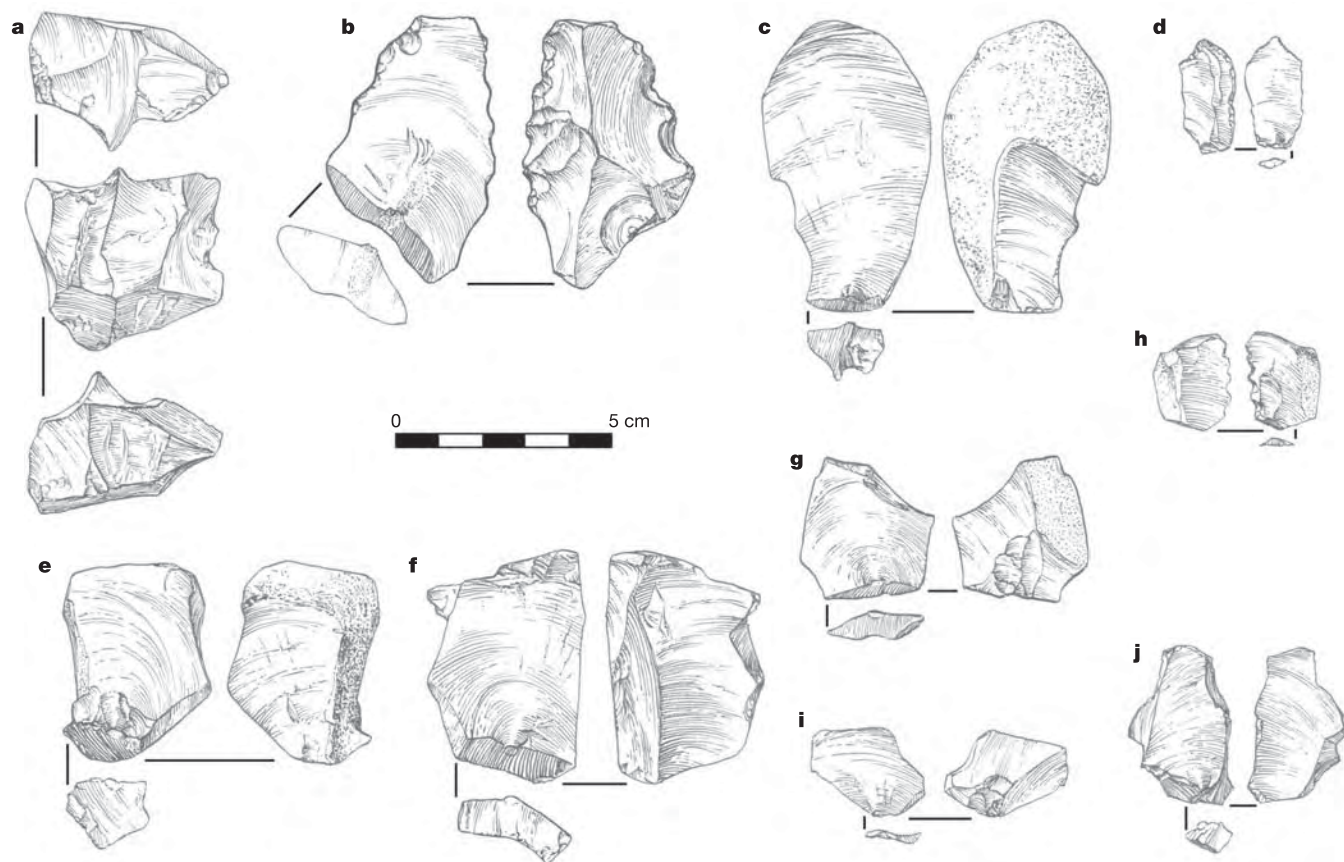


Figure 2 | Lower Palaeolithic flint artefacts from the Cromer Forest-bed Formation at Pakefield. **a**, Core, partly alternate hard-hammer flaking, with several incipient cones of percussion on platforms. **b**, Retouched flake.

c–j, Hard hammer-struck flakes, with previous removals on dorsal surface, often from the same direction. **b**, 'Rootlet bed' (PaB); **c**, laminated silts (PaCi); **a**, **d–j**, 'Unio-bed' (PaCi).

remains include several thermophilous species no longer living in Britain, and the presence of *Hippopotamus* and frost-sensitive insects and plants implies warmer summers and mild winters. Mutual Climatic Range (MCR)¹⁹ analysis of the beetle assemblage enables quantitative estimates to be made of the thermal climate at that time. These suggest that the mean temperature of the warmest month (July) was between 18 and 23 °C and the mean temperature of the coldest months (January/February) was between -6 and +4 °C. Pedogenic carbonate nodules in the 'Rootlet bed' indicate an annual moisture deficit, whereas their isotopic composition reflects intense soil moisture evaporation during their formation²⁰. The combination of warmer summer and winter temperatures together with a strongly seasonal precipitation regime is indicative of a warm, seasonally dry Mediterranean climate.

Both the insects and plants indicate the presence of marshy ground with extensive reedy vegetation and alder trees adjacent to a meandering river with shallow riffles and deeper pools. Oak woodland grew on drier ground with open grassland nearby. This mixture of habitats supported a variety of large browsing and grazing mammals

dominated by *Mammuthus trogontherii*, *Stephanorhinus hundsheimensis*, *Megaloceros savini*, *M. dawkinsi* and *Bison cf. schoetensacki*, and their predators and scavengers (*Homotherium* sp., *Panthera leo*, *Canis lupus* (small) and *Crocuta crocuta*). The floodplain would have provided a resource-rich environment for early humans, with a range of plant and animal resources. An additional attraction, in an area where good quality flint would have been scarce, was the flint-rich river gravels, which provided the raw material for tool manufacture.

Pollen analysis indicates that the channel infill accumulated during an interglacial with regional vegetation dominated by broad-leaf woodland that included *Carpinus* (hornbeam) in its later part¹¹. These sediments were originally correlated with those at the Cromerian stratotype at West Runton, Norfolk, 60 km to the north-west, on the basis of palynology¹¹ and malacology (for example the presence of the extinct freshwater mollusc *Valvata goldfussiana*¹²). This suggestion is broadly supported by the extent of racemization and the pattern of amino acid decomposition in the intracrystalline fraction of *Bithynia opercula* (see Methods and Supplementary Information). However, although these sites are close in age, several

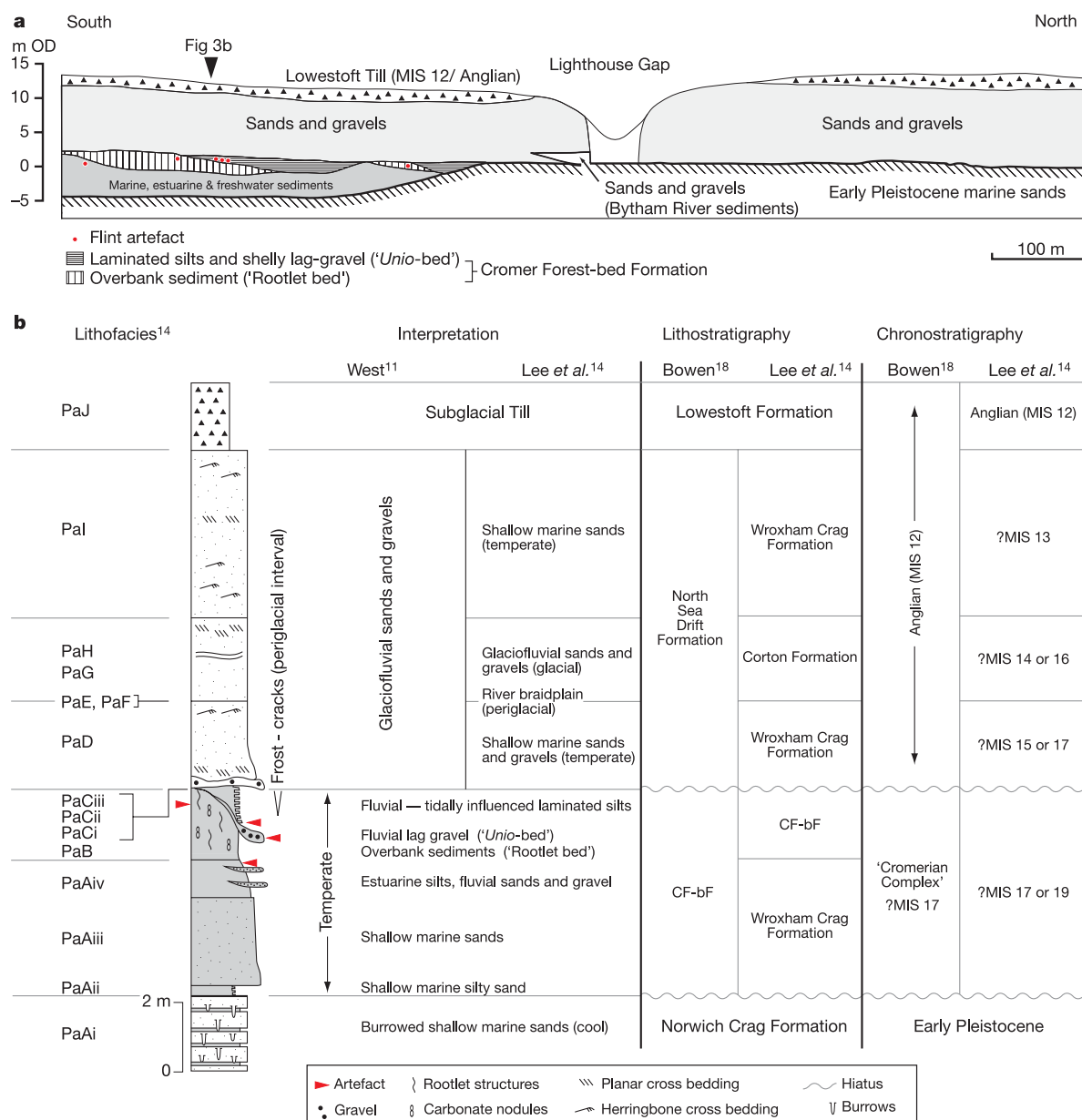


Figure 3 | Stratigraphical context of the Pakefield artefacts. **a**, Coastal sections between Pakefield and Kessingland. **b**, Geological profile and interpretation.

lines of evidence indicate that they are not exactly contemporary. West Runton lacks the important component of southern thermophile plants and beetles seen at Pakefield. Moreover, several large mammals, such as *Hippopotamus*, *Megaloceros dawkinsi* and *Palaeoloxodon antiquus* (the latter recorded only from unstratified material) known from Pakefield/Kessingland have never been found at West Runton, despite the much more extensive collections from that site¹⁶. Furthermore, the occurrence at Pakefield of two vole species of the genus *Mimomys*, *M. savini* and *M. aff. pusillus* (see Supplementary Information), of which only the former is known from West Runton, is also consistent with a difference in age. *M. pusillus* is known from the Early Pleistocene to its latest occurrence in the early Middle Pleistocene Ilynyan Complex (European Russia)²¹. As the latter is overlain by the Don Till (Donian), correlated with MIS 16 (ref. 21), this suggests a minimum age for the Pakefield 'Unio-bed'.

This evidence indicates that sediments containing the artefacts belong to an interglacial in the early part of the 'Cromerian Complex'¹⁰. This interpretation is supported by independent evidence derived from the lithostratigraphy. A maximum age is indicated by palaeomagnetic data from the laminated silts, which show normal polarity¹⁴, consistent with the early part of the Brunhes Chron. A minimum age is indicated by the overlying Lowestoft Till Member, conventionally thought to have been emplaced during MIS 12 (refs 13, 18). The sediments between the CF-bF and the till have been traditionally interpreted as being of glaciofluvial origin, deposited during the Anglian Stage^{11,17}. A longer chronology has recently been proposed on the basis of sedimentological evidence suggesting that the CF-bF may be separated from the MIS 12 till by two separate high sea-stands and two cold episodes¹⁴ (Fig. 3b). By using the premise that temperate-climate marine deposits and cold-climate aggradations in the lower parts of large, temperate latitude rivers can be matched, respectively, to the peaks and troughs of orbitally tuned MIS cycles²², the archaeology can be dated to MIS 17 (about 680 kyr ago) at the very youngest. If evidence for ice-sheet extension across eastern England during MIS 16 is valid²³, then an additional temperate/cold cycle is required and the archaeology at Pakefield could be as old as the later part of MIS 19 (about 750 kyr ago).

The oldest human fossils in northwest Europe are from Mauer⁶, Germany and Boxgrove²⁴, UK, where they are part of vertebrate assemblages that include the water vole *Arvicola terrestris cantiana*. All other northwestern European sites containing early Middle Pleistocene archaeology in association with rich small-mammal assemblages (for example, Miesenheim I in Germany, and Westbury-sub-Mendip and Waverley Wood in the UK)^{6,17,25}, have similarly yielded *Arvicola* (with unrooted molars) rather than *Mimomys* (its ancestor with rooted molars). These early Middle Pleistocene archaeological sites with *Arvicola* have all been correlated with MIS 13, the basis for the belief in a short chronology for human occupation in this region⁶. The discovery at Pakefield of unequivocal artefacts in beds yielding *Mimomys* demonstrates a much longer human occupation of northwest Europe, pre-dating other evidence by as much as 200 kyr. There has been much discussion about what additional social, technological or physiological adaptations humans would have required to colonize northwest Europe compared with their occupation farther south^{5,26}. The Mediterranean climate reconstructed for the archaeological levels at Pakefield suggests that these pioneers were able to spread northwards in familiar climatic conditions, using their existing adaptations.

METHODS

A new technique of amino acid analysis has been developed for geochronological purposes²⁷ that combines a new reverse-phase-high-pressure liquid chromatography (RP-HPLC) method of analysis²⁸ with the isolation of an intracrystalline fraction of amino acids by treatment with bleach²⁹. This combination of techniques results in the analysis of D/L values of multiple amino acids from the chemically protected protein within the biomineral,

enabling both smaller sample sizes and increased reliability of the analysis. The calcitic structure of the opercula of *Bithynia* (a genus of freshwater prosobranch gastropods) has been found to provide the most robust repository for the amino acids yet studied.

Amino acid racemization analyses were undertaken with these procedures on three individual samples of *Bithynia trocheli* opercula (NEaar 1711–1712, 1968) from the laminated silts (PaCii) and four samples of *B. trocheli* opercula (NEaar 2146–2167) from the Unio-bed (PaCi). Each operculum was powdered and bleached for 48 h with 12% NaOCl. Two subsamples were taken: one fraction was directly demineralized and the free amino acids were analysed (referred to as the 'free' fraction, F), and the second was treated with 7 M HCl under N₂ at 110 °C for 24 h (referred to as the 'hydrolysed' fraction, H*) to release the peptide-bound amino acids, thus yielding the 'total' amino acid concentration. Samples were then dried with a centrifugal evaporator and rehydrated for RP-HPLC analysis with 0.01 mM L-homo-arginine as an internal standard.

The amino acid compositions were analysed in duplicate by RP-HPLC using fluorescence detection, following the method of ref. 28. A 2 µl sample was injected and mixed online with 2.2 µl of derivitizing reagent (260 mM n-iso-L-butyl L-cysteine (IBLC) and 170 mM o-phthalaldehyde (OPA) in 1 M potassium borate buffer, adjusted to pH 10.4 with potassium hydroxide pellets). The amino acids were separated on a C₁₈ HyperSil BDS column (5 mm × 250 mm) at 25 °C with a gradient elution of three solvents: sodium acetate buffer (solvent A; 23 mM sodium acetate trihydrate, 1.5 mM sodium azide, 1.3 µM EDTA, adjusted to pH 6.00 ± 0.01 with 10% acetic acid and sodium hydroxide), methanol (solvent C) and acetonitrile (solvent D). The L and D isomers of ten amino acids were routinely analysed. During preparative hydrolysis both asparagine and glutamine undergo rapid irreversible deamination to aspartic acid and glutamic acid, respectively²⁹. It is therefore not possible to distinguish between the acidic amino acids and their derivatives; they are therefore reported together as Asx and Glx.

On the basis of protein decomposition (racemization and degradation of amino acids) we conclude that the results are consistent with attribution to an early part of the 'Cromerian Complex'²⁷. All opercula from Pakefield are significantly older (95% confidence) than those analysed from Waverley Wood, but they cannot be statistically separated from opercula from the type Cromerian West Runton Freshwater Bed. One sample from the Unio-bed (NEaar 2166) seems to be slightly younger than the others. The amino acid data therefore indicate that the Pakefield opercula are at least as old as those from West Runton (see Supplementary Table 2).

Received 3 June; accepted 12 September 2005.

- Dennell, R. Dispersal and colonisation, long and short chronologies: how continuous is the Early Pleistocene record for hominids outside Africa? *J. Hum. Evol.* **45**, 421–440 (2003).
- Gabunia, L. et al. Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: Taxonomy, geological setting, and age. *Science* **288**, 1019–1025 (2000).
- Carbonell, E. et al. Lower Pleistocene hominids and artifacts from Atapuerca-TD6 (Spain). *Science* **269**, 826–829 (1995).
- Manzi, G. Human evolution at the Matuyama–Brunhes boundary. *Evol. Anthropol.* **13**, 11–24 (2004).
- Gamble, C. *The Palaeolithic Societies of Europe* (Cambridge Univ. Press, Cambridge, 1999).
- Dennell, R. & Roebroeks, W. The earliest colonization of Europe: the short chronology revisited. *Antiquity* **70**, 535–542 (1996).
- Newton, E. T. *The Vertebrata of the Forest Bed Series of Norfolk and Suffolk* (Memoirs of the Geological Survey, United Kingdom, London, 1882).
- Reid, C. *The Geology of the Country around Cromer* (Memoirs of the Geological Survey, England and Wales, London, 1882).
- Stuart, A. J. in *The early Middle Pleistocene in Europe* (ed. Turner, C.) 9–24 (Balkema, Rotterdam, 1996).
- Turner, C. in *The early Middle Pleistocene in Europe* (ed. Turner, C.) 295–317 (Balkema, Rotterdam, 1996).
- West, R. G. *The pre-glacial Pleistocene of the Norfolk and Suffolk Coasts* (Cambridge Univ. Press, Cambridge, 1980).
- Preece, R. C. Molluscan evidence for differentiation of interglacials within the 'Cromerian Complex'. *Quat. Sci. Rev.* **20**, 1643–1656 (2001).
- Rose, J., Moorlock, B. S. P. & Hamblin, R. J. O. Pre-Anglian fluvial and coastal deposits in Eastern England: lithostratigraphy and palaeoenvironments. *Quat. Int.* **79**, 5–22 (2001).
- Lee, J. R., Rose, J., Candy, I. & Barendregt, R. Sea-level changes, river activity, soil development and glaciation around the western margins of the southern North Sea basin during the Early and early Middle Pleistocene: evidence from Pakefield, Suffolk, UK. *J. Quat. Sci.* (in the press).
- Blake, J. H. *The Geology of the Country near Yarmouth and Lowestoft*. (Memoirs of the Geological Survey, England and Wales, London, 1890, relating to Horizontal Section 128).

16. Stuart, A. J. & Lister, A. M. The mammalian faunas of Pakefield/Kessingland and Corton, Suffolk, U.K.: evidence for a new temperate episode in the British early Middle Pleistocene. *Quat. Sci. Rev.* **20**, 1677–1692 (2001).
17. Wymer, J. *The Lower Palaeolithic Occupation of Britain* (Wessex Archaeology & English Heritage, Salisbury, 1999).
18. Bowen, D. Q. (ed.) *A New Correlation of Quaternary Deposits in the British Isles* (Geol. Soc. Spec. Rep. 23, London, 1999).
19. Atkinson, T. C., Briffa, K. A. & Coope, G. R. Seasonal temperatures in Britain during the past 22,000 years, reconstructed using beetle remains. *Nature* **325**, 587–592 (1987).
20. Candy, I., Rose, J. & Lee, J. R. A seasonally 'dry' interglacial climate in eastern England during the early Middle Pleistocene: palaeopedological and stable isotopic evidence from Pakefield, UK. *Boreas* (in the press).
21. Pevzner, M., Vangengeim, E. & Tesakov, A. Quaternary subdivisions of Eastern Europe based on vole evolution. *Boll. Soc. Paleont. Ital.* **40**, 269–274 (2001).
22. Bassinot, F. C. et al. The astronomical theory of climate and the age of the Brunhes–Matuyama magnetic reversal. *Earth Planet. Sci. Lett.* **126**, 91–108 (1994).
23. Lee, J. R., Rose, J., Hamblin, R. J. O. & Moorlock, B. S. P. Dating the earliest lowland glaciation of eastern England: a pre-MIS 12 early Middle Pleistocene glaciation. *Quat. Sci. Rev.* **23**, 1551–1566 (2004).
24. Roberts, M. B. & Parfitt, S. A. *Boxgrove: A Middle Pleistocene Hominid Site at Earham Quarry, Boxgrove, West Sussex*. (English Heritage, London, 1999).
25. Bishop, M. J. Earliest record of Man's presence in Britain. *Nature* **253**, 95–97 (1975).
26. Roebroeks, W. Hominid behaviour and the earliest occupation of Europe: an exploration. *J. Hum. Evol.* **41**, 437–461 (2001).
27. Penkman, K. E. H. *Amino Acid Geochronology: A Closed System Approach to Test and Refine the UK Model*. Thesis, Univ. Newcastle (2005).
28. Kaufman, D. S. & Manley, W. F. A new procedure for determining DL amino acid ratios in fossils using reverse phase liquid chromatography. *Quat. Sci. Rev.* **17**, 987–1000 (1998).
29. Sykes, G. A., Collins, M. J. & Walton, D. I. The significance of a geochemically isolated intracrystalline organic fraction within biominerals. *Org. Geochem.* **23**, 1039–1065 (1995).
30. Gibbard, P. L. The history of the great northwest European rivers during the past three million years. *Phil. Trans. R. Soc. Lond. B* **318**, 559–602 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. M. Ashton, A. P. Currant, P. L. Gibbard, N. R. Larkin, M. D. Lewis, S. G. Lewis, W. Roebroeks, J. R. Stewart, C. Turner and A. Charlton for help in the field and/or discussions; T. Lawson, who found the first putative artefact in 2000; and P. Crabb (NHM photo unit) for photographs of the vole molars. A. W. Gentry commented on the bovid remains. R.W.B. acknowledges support from NSERC (Canada) and laboratory assistance provided by R. Enkin and J. Baker at the Geological Survey of Canada-Pacific. J.R.L. publishes with the permission of the Executive Director, British Geological Survey (NERC). We thank the Leverhulme Trust, which supported this work with grants to the Ancient Human Occupation of Britain (AHOB) project (http://www.nhm.ac.uk/hosted_sites/ahob/), and the project 'Mammoth, Vertebrate Fauna and Palaeoenvironment of the Cromerian Stratotype' (A.M.L. and A.J.S.). Amino acid geochronology was funded by English Heritage (Aggregates Levy Sustainability Fund) and NERC. This work was undertaken while J.R.L. was in receipt of a NERC studentship at Royal Holloway, University of London, and the British Geological Survey.

Author Contributions S.A.P., J.R. and A.J.S. coordinated research on the site, and S.A.P. wrote the article, with major contributions from R.C.P., A.J.S., J.R. and C.B.S. The stratigraphical and lithological sequence was described by J.R.L. and J.R., and I.C. investigated the sediment geochemistry. R.M. and P.D. recovered most of the artefacts and mammalian fauna; the artefacts were figured and described by J.J.W. The large mammals were studied by M.B., A.J.S. and A.M.L., and S.A.P. analysed the small mammal fauna. K.E.H.P. and M.J.C. are responsible for the amino acid geochronology and R.W.B. undertook the palaeomagnetic investigations. Beetles were analysed by G.R.C., molluscs by R.C.P., plant macrofossils by M.H.F., and ostracods and foraminifera by J.E.W. R.S. made major contributions in the field and in drawing the figures.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.J.S. (tony@megaafauna.freeseerve.co.uk).

A RhoGDP dissociation inhibitor spatially regulates growth in root hair cells

Rachel J. Carol¹†, Seiji Takeda¹, Paul Linstead¹, Marcus C. Durrant²†, Hana Kakesova^{3,4}, Paul Derbyshire¹, Sinéad Drea¹†, Viktor Zarsky^{3,4} & Liam Dolan¹

Root hairs are cellular protuberances extending from the root surface into the soil; there they provide access to immobile inorganic ions such as phosphate, which are essential for growth¹. Their cylindrical shape results from a polarized mechanism of cell expansion called tip growth in which elongation is restricted to a small area at the surface of the hair-forming cell (trichoblast) tip^{2–4}. Here we identify proteins that spatially control the sites at which cell growth occurs by isolating *Arabidopsis* mutants (*scn1*) that develop ectopic sites of growth on trichoblasts. We cloned *SCN1* and showed that *SCN1* is a RhoGTPase GDP dissociation inhibitor (RhoGDI) that spatially restricts the sites of growth to a single point on the trichoblast. We showed previously that localized production of reactive oxygen species by RHD2/AtrbohC NADPH oxidase is required for hair growth⁵; here we show that *SCN1*/AtrbohC is a component of the mechanism that focuses RHD2/AtrbohC-catalysed production of reactive oxygen species to hair tips during wild-type development. We propose that the spatial organization of growth in plant cells requires the local RhoGDI-regulated activation of the RHD2/AtrbohC NADPH oxidase.

Cell morphogenesis is the process by which cells acquire shape during development and is central to organogenesis in multicellular organisms. The morphogenesis of root hairs is characterized by the

localization of growth to a small dome at the tip of the hair, a process known as tip growth. The restriction of growth to this point is determined by the localized accumulation of reactive oxygen species (ROS) produced by the RHD2/AtrbohC NADPH oxidase⁵. Hairs on mutants that lack RHD2/AtrbohC-derived ROS do not elongate. Nevertheless, treatment of *rhdl2* mutants with ROS restores growth but this growth lacks spatial control. We therefore predict that proteins that spatially regulate plant cell growth will also spatially regulate the activity of the RHD2/AtrbohC NADPH oxidase.

To identify spatial regulators of cellular growth we screened for mutants with delocalized expansion on the outer face of trichoblasts (cells that form root hairs). Two mutants were identified and shown to be allelic to *supercentipede1* (*scn1*) and were designated *scn1-2* and *scn1-3* (ref. 6). There are two aspects to the spatially deregulated growth phenotype of these mutants (Fig. 1b). First, instead of initiating a single growth site along the outer face of the cell (Fig. 1a, e), multiple sites (bulges) are initiated on *scn1* mutant trichoblasts (Fig. 1b–d, f). This indicates that the wild-type (WT) function of *SCN1* is to restrict growth on the outer face to a single site. Second, once initiated, *scn1* mutant hairs develop multiple tip-growing axes from single bulges (Fig. 1d), indicating that *SCN1* also regulates the sites of outgrowth on bulges. The number of growth sites along the outer face of the hair cell in the *scn1-3* mutant is

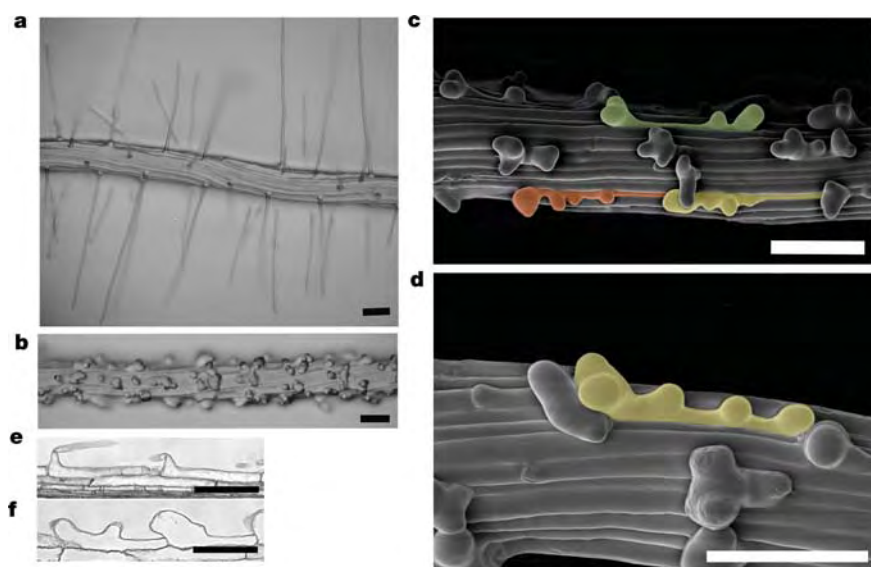


Figure 1 | Phenotype of *scn1* roots. Roots of 5-day-old seedlings. a, WT root hairs. b, *scn1* mutants with short hairs showing multiple sites of hair growth emerging from a trichoblast (hair-forming cell). c, Scanning electron micrograph of root surface showing sites of ectopic growth on the outer face of hair cells. Some hair cells (highlighted with a single colour) develop multiple site of growth along their entire length. Other root hair cells (uncoloured outgrowths) develop multiple sites of growth emerging from enlarged swellings on the outer face of the hair cell. d, Image of a single cell with four sites of outgrowth along its length. e, f, Longitudinal sections through hair cells of WT root (e) and *scn1* root showing extensive outgrowth from the outer face of the trichoblast (f). Scale bars, 60 μ m (a–d) and 50 μ m (e, f).

¹Department of Cell and Developmental Biology, and ²Computational Biology Group, John Innes Centre, Norwich NR4 7UH, UK. ³Department of Plant Physiology, Charles University, Vinicna 5, Prague 2, 128 00, Czech Republic. ⁴Institute of Experimental Botany, Academy of Sciences of the Czech Republic, Rozvojova 135, Prague 6, 165 02, Czech Republic. †Present addresses: École Pratique des Hautes Études, 46 rue de Lille, 75007 Paris, France (R.J.C.); School of Applied Sciences, Ellison Building, University of Northumbria, Newcastle upon Tyne NE1 8ST, UK (M.C.D.); Department of Molecular, Cellular and Developmental Biology, PO Box 208104, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8104, USA (S.D.).

3.1 ± 0.1 (mean $\pm$ s.e.m.), $n = 88$ cells), whereas WT cells develop single hairs per cell⁶. Together these data indicate that *SCN1* controls the sites at which growth occurs during root hair development.

We showed previously that treating *rhd2* mutant roots with ROS restored root hair growth but that this growth was spatially deregulated⁵. This indicates that the local production of ROS might have been at least partly responsible for the localization of cell growth. To test whether *SCN1* controls the site of outgrowth by spatially controlling ROS production, we revealed apoplastic ROS production in WT and *scn1* hairs with the use of nitroblue tetrazolium (NBT). In the presence of superoxide, NBT is reduced to a blue formazan deposit⁷. ROS are produced on the surface of the bulge as hair formation begins in both WT and *scn1* (Fig. 2) roots. In WT roots, ROS become tightly focused to the tip during growth (Fig. 2a, f–h) and disappear once tip growth ceases (Supplementary Fig. 1). By contrast, during hair growth in *scn1* mutants, ROS are produced in ectopic foci and over a greater area of the cell surface, encompassing sites of outgrowth (Fig. 2c, i, j). This spatial deregulation of ROS production in plants that lack *SCN1* activity indicates that *SCN1* might promote the formation of the single focus of ROS production in WT roots, which in turn promotes growth along a single axis. To verify that the NBT staining in the root hair reports the activity of RHD2/AtrbohC and not another ROS-producing enzyme, *rhd2* mutants were incubated with NBT. No blue NBT staining was observed in the epidermis of *rhd2* roots, indicating that NBT staining of hairs reports only RHD2/AtrbohC activity (Fig. 2e). Moreover, to show that a haem-containing oxidoreductase such as RHD2/AtrbohC is required for the NBT staining, an inhibitor of this class of enzyme, diphenylene iodonium (DPI), was added at the same time as NBT⁸. No NBT staining of hairs was observed in either WT or *scn1* roots when incubated in DPI, indicating that *SCN1* acts by deregulating the activity of a haem-containing oxidoreductase (Fig. 2b, d). Together these data indicate that spatial deregulation of RHD2/AtrbohC is responsible for the accumulation of delocalized ROS in *scn1* mutants, and therefore that *SCN1* spatially regulates RHD2/AtrbohC activity.

To verify genetically that *SCN1* controls RHD2/AtrbohC activity during hair growth, we characterized the phenotype of *scn1 rhd2* double mutants. *scn1* mutants develop ectopic sites of growth on the outer face of the hair cell and multiple hairs from single initiation sites. Both phenotypes are suppressed in the *scn1 rhd2* double mutant (Fig. 3a). Single initiation sites develop on the *scn1 rhd2* double mutant at the end of the cell nearest the meristem, from which limited elongation may occasionally occur as observed in the *rhd2* single mutant. Whereas WT and *rhd2* roots generally produce a single growth point per cell, *scn1* roots always produce multiple (two to five) growth points (the mean number of growth points per cell was 3.1 ; Fig. 3b). However, in *scn1 rhd2* roots, 85% of cells make a single growth point (1.2 ± 0.04 growth points per cell (mean $\pm$ s.e.m.); $n = 121$ cells; Fig. 3a, e). Furthermore, the development of ectopic sites of ROS accumulation as revealed by NBT staining evident in *scn1* single mutants is absent from the *scn1 rhd2* double mutant (Fig. 3b–d). This indicates that the ectopic sites of ROS accumulation in the *scn1* mutant are produced by the action of RHD2/AtrbohC. Together these data indicate that RHD2 activity is required to form the multiple growth points observed in the *scn1* mutants. This is consistent with the regulation of RHD2/AtrbohC by *SCN1* during the formation of a single axis of growth in WT hairs, although it does not indicate whether this regulation is direct or indirect.

To determine the molecular basis of *SCN1* action in hair growth, we isolated the *SCN1* gene by map-based cloning (Supplementary Fig. 2a). *SCN1* corresponds to the gene *At3g07880*, which encodes the RhoGTPase GDP dissociation inhibitor AtRhoGDI1 (ref. 9). Hereafter we will refer to the gene as *SCN1/AtRhoGDI1*. *SCN1/AtRhoGDI1* transcripts are present in WT roots, stems, siliques and inflorescences, but absent from leaves (Supplementary Fig. 2c). *In situ* hybridization with a *SCN1/AtRhoGDI1* antisense probe shows strong

expression of the gene in the epidermis of WT roots (Supplementary Fig. 2d), in which the *Scn1*[−] phenotype indicates that the gene is active. Mutations in *scn1-3* and *scn1-1* create premature stop codons in the full-length complementary DNA at Arg 42 (nucleotide 328) and Trp 86 (nucleotide 462), respectively (Supplementary Fig. 2b). No *SCN1/AtRhoGDI1* mRNA is detectable by reverse transcriptase-mediated polymerase chain reaction (RT–PCR) in *scn1-1* and *scn1-3* mutants. *scn1-2* has a G $\rightarrow$ A change at nucleotide 749, leading to the substitution of Glu for Gly 181 (Supplementary Fig. 2b). Mutant mRNA is present in *scn1-2* roots (Supplementary Fig. 2c). To summarize, *SCN1/AtRhoGDI1* is a key spatial regulator of plant cell growth.

It has previously been shown that *SCN1/AtRhoGDI1* interacts with AtROP4 and AtROP6, members of the RhoGTPases of plants (ROP) family¹⁰ in yeast two-hybrid and protein pull-down assays⁹. Several ROPs are expressed in roots, and misexpression studies have indicated that ROPs such as ROP2 might be required for hair elongation^{11,12}. To determine whether the interaction between *SCN1/AtRhoGDI1* and ROPs was required for root hair growth, we examined the interaction between *scn1-2* and ROPs. First we constructed a homology model of the *SCN1/AtRhoGDI1*–AtROP4 complex, using the structure of the human RhoGDI–Rac1 complex as a template¹³ (Supplementary Fig. 3a). The model includes a geranylgeranyl group covalently attached to Cys 193 of ROP4 and located within the internal cavity of *SCN1/AtRhoGDI1*. From the model we found that the Gly $\rightarrow$ Glu mutation of *scn1-2* is located in the protein–protein interface between *SCN1/AtRhoGDI1* and AtROP4. The glutamate side chain of the mutant would project into the interface, producing unfavourable steric and electrostatic contacts with AtROP4. Furthermore, the WT Gly 182 residue occupies a region of the Ramachandran plot that is disallowed for

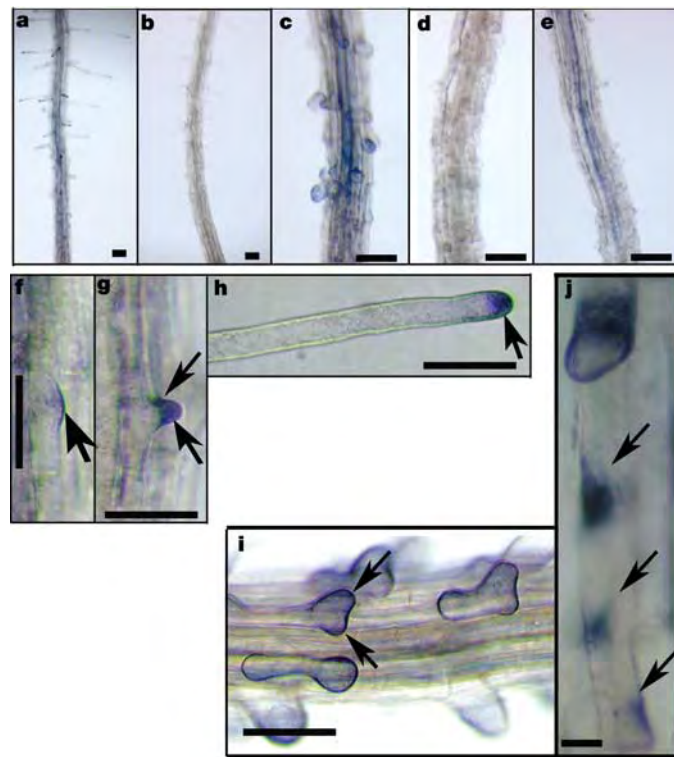


Figure 2 | ROS localization and hair growth. **a**, WT, stained with NBT only. **b**, WT, stained with NBT plus 50 μ M DPI. **c**, *scn1*, stained with NBT only. **d**, *scn1*, stained with NBT plus 50 μ M DPI. **e**, *rhd2*, stained with NBT only. Scale bars, 100 μ m. **f–h**, NBT staining in WT hairs at initiation (**f**), initiation of tip growth (**g**) and during tip growth (**h**). Scale bars, 30 μ m. **i**, NBT staining in *scn1* hairs. Scale bar, 30 μ m. **j**, Three foci (arrows) of NBT staining along the length of a single hair-forming cell. Scale bar, 10 μ m.

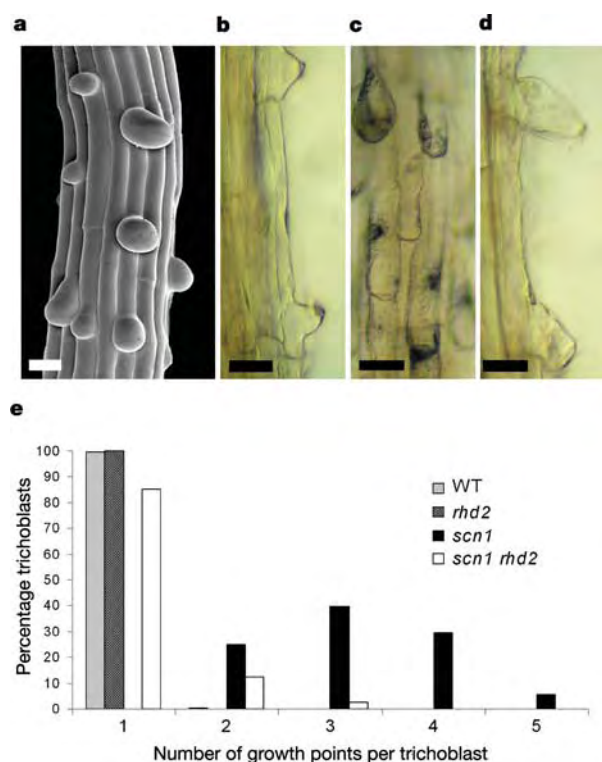


Figure 3 | Phenotype of *scn1 rhd2* double mutant showing that the multiple hair-initiation phenotype characteristic of *scn1* mutant is suppressed in the *rhd2* mutant background. **a**, Cryoscanning electron micrograph of root epidermis of *scn1 rhd2*. Scale bar, 20 μ m. **b–d**, NBT staining of *rhd2* mutant (**b**), *scn1-1* mutant (**c**) and *scn1-1 rhd2* double mutant (**d**). Scale bar, 15 μ m. **e**, Frequency distribution of the number of growth points per hair cell observed in cryoscanning electron micrographs. *rhd2*, $n = 78$ hair cells (8 roots); *scn1*, $n = 88$ hair cells (9 roots); *scn1 rhd2*, $n = 121$ hair cells (11 roots).

other amino acids. This implies that in addition to the direct effects of the Glu side chain, there may also be some perturbation of the AtRhoGDI backbone in this region, further weakening its interaction with the ROP. The WT glycine residue that is replaced in *scn1-2* is conserved in all GDIs described so far^{9,14}. To determine whether the interaction between ROP and SCN1/AtRhoGDI is defective in *scn1-2* mutants we compared the ROP-binding ability of *scn1-2* and WT protein. The pull-downs indicate that 60% less ROP binds to *scn1-2* than to SCN1/AtRhoGDI1 (WT) (Supplementary Fig. 3b–d). Because the decrease in ROP binding observed in the *scn1-2* mutant results in a phenotype indistinguishable from that observed in *scn1* complete loss-of-function mutations (*scn1-1* and *scn1-3*), it indicates that the SCN1/AtRhoGDI-regulated ROP activity is required for root hair development. Consistent with this conclusion is the observation that ROP2::YFP is mislocalized in *scn1* mutants. In WT roots, ROP2 is present at the cell surface upon hair initiation, and as elongation proceeds it spreads to the cytoplasm in the vicinity of the growing tip¹⁵ (Supplementary Fig. 4a–d). In *scn1* mutants, ROP2 accumulates at the cell surface in ectopic locations where supernumerary hair initiation sites are being initiated (Supplementary Fig. 4e, f) and it remains at the cell surface subsequently during elongation (Supplementary Fig. 4g, h). This indicates that the subcellular localization of ROP2 might be regulated by SCN1. Together these data indicate that SCN1/RhoGDI spatially controls growth in root hairs through its regulation of ROP proteins. Further evidence supporting a role for a RhoGDI and a ROP during plant cell growth comes from the observation that overexpression of *SCN1/AtRhoGDI1* suppresses the defective growth caused by overexpression of AtROP1 in tobacco pollen tubes¹⁶.

Thus, SCN1/AtRhoGDI spatially regulates the pattern of growth in hair cells. We propose that it does this by regulating the activity of a ROP GTPase and the focused production of RHD2/AtRbohC-derived ROS at the sites of growth. Because small GTPases related to ROPs control the activity of mammalian NADPH oxidases similar to RHD2/AtRbohC^{17,18}, it is possible that a ROP regulates RHD2/AtRbohC directly. Nevertheless, we cannot completely rule out the possibility that SCN1/AtRhoGDI1 spatially regulates the production of ROS by other mechanisms, for example through a role in regulating microfilament organization or endocytosis. SCN1/AtRhoGDI1 is therefore a key spatial regulator of growth in plant cells. Given that RHD2/AtRbohC-derived ROS are required for diffuse growth observed in other cell types in the plant, we predict that SCN1/AtRhoGDI1 might also have a function in spatially regulating the cell faces on which growth occurs.

METHODS

Plants. *Arabidopsis thaliana* (L.) Heynh plants were grown as described previously⁵. Ethane methyl sulphonate (EMS) mutagenesis of Columbia (Col-0) ecotype and screening M₂ seedlings for root-hair-morphology mutants have been described previously¹⁹. The two new mutants segregated in a recessive, nuclear mendelian manner. Allelism tests were made between these mutants and *scn1-1* (ref. 6) and all of the F₁ progeny of reciprocal crosses had an *Scn1*[−] phenotype. *scn1-3*, used for phenotypic characterization, was backcrossed four times to Columbia. *scn1-3* and *rhd2-1* were crossed. The F₁ generation all had a WT phenotype. The F₂ progeny segregated with a new phenotype. The phenotype of the putative double mutant was verified by crossing to *rhd2-1* and *scn1-3* parents. The F₁ progeny of test crosses all had *Rhd2*[−] or all had *Scn1*[−] phenotypes, respectively.

Isolation of SCN1 gene. Genomic DNA was isolated from WT and *scn1* mutants with Qiagen plant genomic preparations. Genomic DNA from the *At3g07880/AtRhoGDI1* locus was amplified with primers A and B and sequenced using primers A, B, C, D and E with duplicate samples (primer sequences are shown in Supplementary Fig. 2). Note that sequencing of the cDNA and comparison with full-length cDNA clones from the *Arabidopsis thaliana* database (<http://www.arabidopsis.org>) indicates that the ATG start codon might be further upstream than described previously⁹. Primers A and B were used to amplify *SCN1/AtRhoGDI1* cDNA. The APT1 primers and RT-PCR methods were described previously²⁰. Whole-mount *in situ* hybridization with riboprobes was used to determine the cells in which the SCN1/AtRhoGDI1 is expressed. The protocol was a modification of that described in ref. 21 (details in Supplementary Methods).

Imaging and ROS. Microscopic methods were described previously^{5,19}. Nitro-blue tetrazolium chloride (Roche Diagnostics GmbH) was used to stain for the site of superoxide production⁷. Seedlings (5 days old) grown on solid medium were covered with freshly made NBT solution (0.5 mg ml^{−1} NBT in 0.1 M Tris-HCl, 0.1 M NaCl, 0.05 M MgCl₂ pH 9.5 or 0.5 mg ml^{−1} NBT in 0.1 M potassium phosphate with or without 0.1 M NaCl at pH 7). Plates were left for 2 h at 20 °C for the colour to develop. Roots were imaged under bright-field illumination on a Microphot microscope (Nikon). Diphenylene iodonium chloride (Aldrich) was dissolved in methanol and added in the NBT buffer. The same volume of methanol was added to the control.

Homology modelling. Homology modelling was performed with the Homology module of Insight II, version 2000.1 (Accelrys Ltd). The X-ray crystal structures of homologous proteins for RhoGDI (accession code 1DOA) and ROP4 (accession codes 1I4D and 1HH4) were retrieved from the Protein Data Bank²² and used as templates for building the individual structures. The first 40 residues of RhoGDI and the last 3 residues of ROP4 were omitted. The geranylgeranyl group was attached to ROP4 and the individual protein structures were docked using structure 1HH4 as a template. The structures of the component proteins and the final complex were each energy-minimized to a derivative of 1.0 with the consistent-valence force field within the Discover module, and the model was validated with Procheck.

Pull-down assay. Fresh cauliflower (*Brassica oleracea* var. *botrytis* L.) was ground in liquid nitrogen and extract was prepared from 10 g of plant tissue with 15 ml of extraction buffer (20 mM HEPES pH 6.8, 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 1 mM phenylmethyl sulphonyl fluoride). Extract was filtered through two layers of Miracloth and then centrifuged at 10,000 g for 10 min to prepare crude cell extract with the majority of endomembrane vesicles in the supernatant. Extract (1 ml) was incubated for 1 h at 4 °C with 10 μ l of glutathione S-transferase (GST)-tagged SCN1/AtRhoGDI1; another 1 ml of the extract was incubated with 10 μ l of GST-*scn1-2/AtRhoGDI1*. As a control, free GST was

used. The resins were washed three times with PBS buffer and the bound proteins were analysed by 10% SDS-PAGE. They were transferred to nitrocellulose membrane by semi-dry blot (Biometra), then stained briefly with Ponceau to reveal proteins after blotting. Membrane was blocked overnight at 4 °C in TBS containing 0.5% Tween 20 and 5% milk powder (blocking solution), washed once for 10 min with TBS containing 0.5% Tween 20 and incubated for 1 h at room temperature with anti-NtRop1 antibody (see Supplementary Methods) diluted 1:1,000, washed three times (10 min each) with TBS containing 0.5% Tween 20, incubated for 1 h at room temperature with horseradish peroxidase-conjugated secondary anti-rabbit IgG antibody (Promega) diluted 1:2,500, and finally washed three times for 10 min with TBS containing 0.5% Tween 20 and developed with a chemoluminescence procedure.

Received 24 March; accepted 7 September 2005.

- Gahoonia, T. S., Care, D. & Nielsen, N. E. Root hairs and phosphorus acquisition of wheat and barley cultivars. *Plant Soil* **191**, 181–188 (1997).
- Dolan, L. *et al.* Clonal relationships and cell patterning in the root epidermis of *Arabidopsis*. *Development* **120**, 2465–2475 (1994).
- Shaw, S. L., Dumais, J. & Long, S. R. Cell surface expansion in polarly growing root hairs of *Medicago truncatula*. *Plant Physiol.* **124**, 959–970 (2000).
- Dumais, J., Long, S. R. & Shaw, S. L. The mechanics of surface expansion anisotropy in *Medicago truncatula* root hairs. *Plant Physiol.* **136**, 3266–3275 (2004).
- Foreman, J. *et al.* Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature* **422**, 442–446 (2003).
- Parker, J. S., Cavell, A. C., Dolan, L., Roberts, K. & Grierson, C. S. Genetic interactions during root hair morphogenesis in *Arabidopsis*. *Plant Cell* **12**, 1961–1974 (2000).
- Fryer, M. J., Oxborough, K., Mullineaux, P. M. & Baker, N. R. Imaging of photo-oxidative stress responses in leaves. *J. Exp. Bot.* **53**, 1249–1254 (2002).
- Barcelo, A. R. Some properties of the H₂O₂/O₂-generating system from the lignifying xylem of *Zinnia elegans*. *Free Radical Res.* **31** (Suppl.), S147–S154 (1999).
- Bischoff, F., Vahlkamp, L., Molendijk, A. & Palme, K. Localization of AtROP4 and AtROP6 and interaction with the guanine nucleotide dissociation inhibitor AtRhoGDI1 from *Arabidopsis*. *Plant Mol. Biol.* **42**, 515–530 (2000).
- Vernoud, V., Horton, A. C., Yang, Z. & Nielsen, E. Analysis of the small GTPase gene superfamily of *Arabidopsis*. *Plant Physiol.* **131**, 1191–1208 (2003).
- Jones, M. A. *et al.* The *Arabidopsis* Rop2 GTPase is a positive regulator of both root hair initiation and tip growth. *Plant Cell* **14**, 763–776 (2002).
- Molendijk, A. J. *et al.* *Arabidopsis thaliana* Rop GTPases are localized to tips of root hairs and control polar growth. *EMBO J.* **20**, 2779–2788 (2001).
- Grizot, S. *et al.* Crystal structure of the Rac1–RhoGDI complex involved in NADPH oxidase activation. *Biochemistry* **40**, 10007–10013 (2001).
- Rivero, F. *et al.* Defects in cytokinesis, actin reorganization and the contractile vacuole in cells deficient in RhoGDI. *EMBO J.* **21**, 4539–4549 (2002).
- Xu, J. & Scheres, B. Dissection of *Arabidopsis* ADP-RIBOSYLATION FACTOR 1 function in epidermal cell polarity. *Plant Cell* **17**, 525–536 (2005).
- Fu, Y., Wu, G. & Yang, Z. Rop GTPase-dependent dynamics of tip-localized F-actin controls tip growth in pollen tubes. *J. Cell Biol.* **152**, 1019–1032 (2001).
- Segal, A. W. & Abo, A. The biochemical basis of the NADPH oxidase of phagocytes. *Trends Biochem. Sci.* **18**, 43–47 (1993).
- Abo, A. *et al.* Activation of the NADPH oxidase involves the small GTP-binding protein p21rac1. *Nature* **353**, 668–670 (1991).
- Favery, B. *et al.* KOJAK encodes a cellulose synthase-like protein required for root hair cell morphogenesis in *Arabidopsis*. *Genes Dev.* **15**, 79–89 (2001).
- Cowling, R. J., Kamiya, Y., Seto, H. & Harberd, N. P. Gibberellin dose-response regulation of GA4 gene transcript levels in *Arabidopsis*. *Plant Physiol.* **117**, 1195–1203 (1998).
- Coen, E. S. *et al.* *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**, 1311–1322 (1990).
- Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Cvrckova for the anti-Rop antibody, C. S. Grierson for providing *scn1-1* seeds, and B. Scheres and J. Xu for providing the ROP2::YFP fusion construct and ROP2::YFP-transformed seed. S.T. is supported by a grant from the BBSRC to L.D., who is also supported by EU TIPNET. R.C., L.D. and M.D. are funded by a grant-in-aid to the John Innes Centre by the BBSRC. P.D. and S.D. were supported by grants to L.D., P. Shaw and J. Doonan and from Syngenta and the BBSRC. V.Z. and H.K. are funded by EU TIPNET and GACR (Grant Agency of Czech Republic).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.D. (liam.dolan@bbsrc.ac.uk).

BDNF from microglia causes the shift in neuronal anion gradient underlying neuropathic pain

Jeffrey A. M. Coull^{1,2*}, Simon Beggs^{3*}, Dominic Boudreau¹, Dominick Boivin¹, Makoto Tsuda^{3,4}, Kazuhide Inoue⁴, Claude Gravel^{5,6}, Michael W. Salter³ & Yves De Koninck^{1,2,6}

Neuropathic pain that occurs after peripheral nerve injury depends on the hyperexcitability of neurons in the dorsal horn of the spinal cord^{1–3}. Spinal microglia stimulated by ATP contribute to tactile allodynia, a highly debilitating symptom of pain induced by nerve injury⁴. Signalling between microglia and neurons is therefore an essential link in neuropathic pain transmission, but how this signalling occurs is unknown. Here we show that ATP-stimulated microglia cause a depolarizing shift in the anion reversal potential (E_{anion}) in spinal lamina I neurons. This shift inverts the polarity of currents activated by GABA (γ -amino butyric acid), as has been shown to occur after peripheral nerve injury⁵. Applying brain-derived neurotrophic factor (BDNF) mimics the alteration in E_{anion} . Blocking signalling between BDNF and the receptor TrkB reverses the allodynia and the E_{anion} shift that follows both nerve injury and administration of ATP-stimulated microglia. ATP stimulation evokes the release of BDNF from microglia. Preventing BDNF release from microglia by pretreating them with interfering RNA directed against BDNF before ATP stimulation also inhibits the effects of these cells on the withdrawal threshold and E_{anion} . Our results show that ATP-stimulated microglia signal to lamina I neurons, causing a collapse of their transmembrane anion gradient, and that BDNF is a crucial signalling molecule between microglia and neurons. Blocking this microglia–neuron signalling pathway may represent a therapeutic strategy for treating neuropathic pain.

Peripheral nerve injury (PNI)-induced tactile allodynia depends on a depolarizing shift in the E_{anion} of spinal lamina I (LI) neurons in the dorsal spinal cord, causing disinhibition and, in some cases, converting the GABA_A-receptor- and glycine-receptor-mediated inhibition to excitation⁵. As microglia are implicated in the induction of neuropathic pain⁴, we considered that these glia cells may affect E_{anion} in LI neurons. To investigate this possibility, we administered microglia to the lumbar spinal level of naive rats by an intrathecal catheter as described⁴, and subsequently made perforated-patch and whole-cell recordings from LI neurons in acute spinal cord slices prepared from these rats. Prior to slice preparation, we determined the paw withdrawal threshold for each rat^{4,5}. Administering control unstimulated microglia produced no change in paw withdrawal threshold, whereas microglia that had been stimulated with ATP caused a progressive decrease in paw withdrawal threshold over the 5 h of testing after injection (Fig. 1a). Cortically and spinally derived microglia produced a comparable decrease in paw withdrawal threshold. Owing to its larger size, the cortex yielded more microglia and was therefore used for all subsequent experiments.

Electrophysiological recordings were made from slices prepared 5 h after intrathecal microglia administration. Using voltage-clamp recording from LI neurons, we found that in spinal slices taken from rats injected with control microglia E_{anion} was -68.3 ± 1.8 mV (mean $\pm$ s.e.m; $n = 6$; Fig. 1b). By contrast, in LI neurons from rats that had been administered ATP-stimulated microglia, E_{anion} was -61.6 ± 1.1 mV ($n = 16$, $P < 0.0001$). In addition, using current-clamp recordings, we found that the GABA response switched from hyperpolarizing in control rats to depolarizing in rats treated with ATP-stimulated microglia (Fig. 1c).

We reasoned that, to effect the shift in E_{anion} , ATP-stimulated microglia must signal to the LI dorsal horn neurons. Activated microglia secrete various biologically active signalling molecules, one of which, BDNF⁶, has been implicated in the hypersensitivity of dorsal horn neurons that follows sensitization and inflammation^{7–9} and in anion gradient shifts in the hippocampus¹⁰. To test whether BDNF could trigger shifts in pain hypersensitivity and in LI neuronal E_{anion} similar to those resulting from the application of ATP-stimulated microglia, we administered recombinant BDNF intrathecally to naive rats. This locally delivered BDNF produced a significant and transient decrease in paw withdrawal threshold comparable to that produced by ATP-stimulated microglia (Fig. 2a).

To determine whether BDNF could cause a shift in E_{anion} , we bath-applied it to spinal slices taken from naive rats. E_{anion} of LI neurons ($n = 9$) in slices treated with BDNF (>90 min) was significantly less negative than that of LI neurons from control untreated slices ($n = 9$; $P < 0.005$; Fig. 2b). Thus, it seemed possible that responses to GABA might be excitatory, rather than inhibitory, during BDNF administration. We investigated this possibility by monitoring intracellular calcium ($[\text{Ca}^{2+}]_i$) after brief applications of GABA in LI neurons ($n = 96$). During perfusion with BDNF and in the presence of glutamate receptor blockers, the proportion of neurons responding to GABA with a rise in $[\text{Ca}^{2+}]_i$ increased over time, reaching 31% of neurons recorded between 80 and 120 min ($P < 0.05$; Fig. 2c). The rise in $[\text{Ca}^{2+}]_i$ was prevented by the GABA_A receptor blocker bicuculline ($n = 18$; $P < 0.05$), confirming that the effect was mediated by GABA_A receptors. Thus, acute administration of BDNF in slices caused a depolarizing shift in E_{anion} and, in about 30% of the cells, caused GABA to produce net excitation.

To determine the effects of sustained prolonged exposure to BDNF *in vivo*, we administered a BDNF-transducing recombinant adenovirus (adBDNF)¹¹ intrathecally to the rats ($n = 16$). After injection, adBDNF caused a progressive decrease in paw withdrawal threshold over the 4 d of testing. By contrast, a control adenovirus had no effect

¹Division de Neurobiologie Cellulaire, Centre de Recherche Université Laval Robert-Giffard, Québec, Québec G1J 2G3, Canada. ²Department of Pharmacology & Therapeutics, McGill University, Montréal, Québec H3G 1Y6, Canada. ³Programme in Brain and Behaviour, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. ⁴Department of Molecular and System Pharmacology, Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi, Fukuoka 812-8582, Japan. ⁵Division de Neurobiologie Systémique, Centre de recherche Université Laval Robert-Giffard, Québec, Québec G1J 2G3, Canada. ⁶Département de Psychiatrie, Université Laval, Québec, Québec G1K 7P4, Canada.

*These authors contributed equally to this work.

on paw withdrawal threshold over the same period ($n = 6$; $P < 0.005$; Fig. 2d). Because of the prolonged effect of the adBDNF treatment, we could test for changes in E_{anion} in slices taken from treated rats. E_{anion} in LI neurons from adBDNF-injected rats ($n = 7$) was significantly less negative than that in LI neurons from rats treated with the control adenovirus ($n = 4$; $P < 0.01$; Fig. 2e, f). Moreover, GABA application caused some LI neurons from

adBDNF-injected rats to fire action potentials (2/7 cells tested), a feature that was not observed in control conditions. Thus, similar to acute administration of BDNF, sustained local release caused a decrease in paw withdrawal threshold, caused a depolarizing shift in E_{anion} and, in some neurons, could switch the action of GABA from inhibitory to excitatory.

These results show that exogenous BDNF is sufficient to cause tactile allodynia and a shift in E_{anion} , raising the possibility that BDNF might be an endogenous mediator of the sequelae of PNI. To investigate this, we used a function-blocking antibody against the TrkB receptor (anti-TrkB) and a BDNF-sequestering fusion protein (TrkB-Fc), each of which blocks the effects of BDNF^{7,9,12,13}. We administered anti-TrkB or TrkB-Fc intrathecally to rats that had developed allodynia 2 weeks after PNI. Both agents acutely reversed the decrease in paw withdrawal threshold ($n = 7$ and $n = 4$, respectively, $P < 0.05$; Fig. 3a). By contrast, vehicle administration to rats with PNI produced no change in withdrawal threshold (data not shown). To determine whether BDNF–TrkB signalling is necessary for the nerve injury-induced shift in E_{anion} in LI neurons, we examined the effect of anti-TrkB applied acutely to spinal cord slices taken from rats with allodynia 2 weeks after PNI. We found that E_{anion} of LI neurons in slices treated with anti-TrkB ($n = 7$) was significantly more negative than E_{anion} in vehicle-treated slices ($n = 6$; $P < 0.05$; Fig. 3b, c). Together, these findings indicate that endogenous BDNF is necessary to sustain both the tactile allodynia and the depolarizing shift in E_{anion} in LI neurons that result from PNI.

If microglia are the source of BDNF, then interfering with BDNF–TrkB signalling should prevent the tactile allodynia and the shift in LI neuronal E_{anion} produced by administering ATP-stimulated microglia (Fig. 4). We tested the first part of this prediction by delivering ATP-stimulated microglia together with either anti-TrkB or TrkB-Fc; the two cocktails produced no change in paw withdrawal threshold over the 5 h after intrathecal injection ($n = 8$ and $n = 7$, respectively; Fig. 4a). By contrast, allodynia developed progressively after the administration of ATP-stimulated microglia without these agents ($n = 8$). Notwithstanding these results, it remained possible that microglia stimulated with ATP might have provoked the release of BDNF from cells in the spinal cord and that the blockers interfered with the action of BDNF from this source rather than from the administered microglia *per se*.

To differentiate between these two possibilities, it was necessary to suppress BDNF production in the microglia that were subsequently stimulated with ATP and then administered intrathecally. We pre-treated the cultured microglia with double-stranded short interfering RNA directed against BDNF (BDNF siRNA¹⁴). The microglia were then stimulated with ATP and, when injected intrathecally into naive rats, did not cause a change in withdrawal threshold ($n = 7$; Fig. 4a). To control for possible nonspecific effects of siRNA, we treated some microglia with a scrambled version of the BDNF siRNA before ATP stimulation; these microglia elicited a robust allodynia ($n = 4$, Fig. 4a). In addition, ATP-evoked Ca^{2+} responses in microglia treated with anti-TrkB or with BDNF siRNA did not differ from those of vehicle-treated control microglia, showing that anti-TrkB or BDNF siRNA treatment did not affect the response of the microglia to ATP (Fig. 4d). Yet, interfering with BDNF–TrkB signalling prevented microglia-induced tactile allodynia.

We then addressed the prediction that the depolarizing shift in E_{anion} produced by ATP-stimulated microglia could be prevented by interfering with BDNF–TrkB signalling. E_{anion} in LI neurons from rats receiving ATP-stimulated microglia together with anti-TrkB or after BDNF siRNA pretreatment did not differ significantly from that in LI neurons from rats receiving unstimulated microglia. However, E_{anion} in LI neurons from either of these groups of rats was significantly more negative than that in LI neurons from rats that had received ATP-stimulated microglia with vehicle (Fig. 4c). Thus, anti-TrkB and BDNF siRNA prevented the shift in E_{anion} induced by ATP-stimulated microglia. Moreover, ATP stimulation ($n = 3$), but

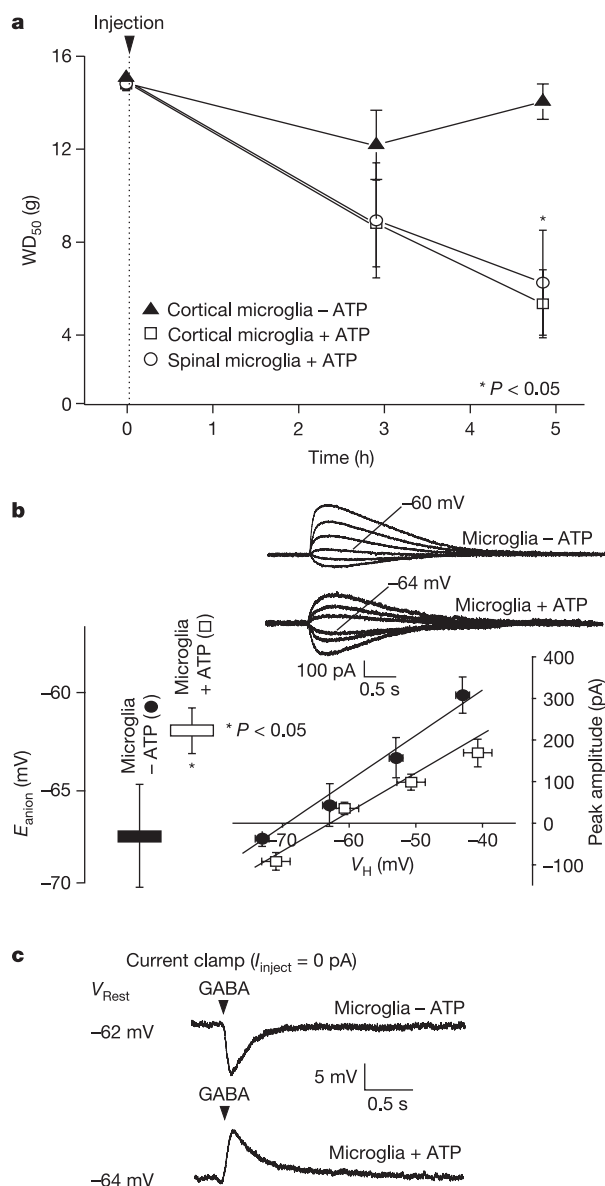


Figure 1 | Spinal delivery of ATP-stimulated microglia to rats evokes allodynia and a depolarizing shift in E_{anion} in spinal LI neurons. **a**, In naive rats, local spinal delivery of ATP-stimulated microglia (of cortical or spinal origin), but not resting microglia, by means of an intrathecal catheter caused a significant decrease in the mean paw withdrawal threshold (WD_{50} ; vertical bars indicate the s.e.m.). **b**, Left, comparison of the mean $\pm$ s.e.m. E_{anion} recorded in LI neurons from rats injected with resting or ATP-stimulated microglia. Note that there were no significant changes in the resting membrane potential of the cells. Right, the mean $\pm$ s.e.m. peak current evoked by GABA, measured in LI neurons at various values of membrane potential (V_m) in slices taken from the rats in **a**. Horizontal error bars represent interneuron differences. Inset, representative raw traces. **c**, Representative traces in current-clamp recording mode showing that, at resting membrane potential, the postsynaptic response to GABA was depolarizing in a LI neuron taken from a rat with a WD_{50} of 3.4 g, in contrast to the response in a LI neuron taken from a rat with a WD_{50} of 12.6 g, where GABA was hyperpolarizing.

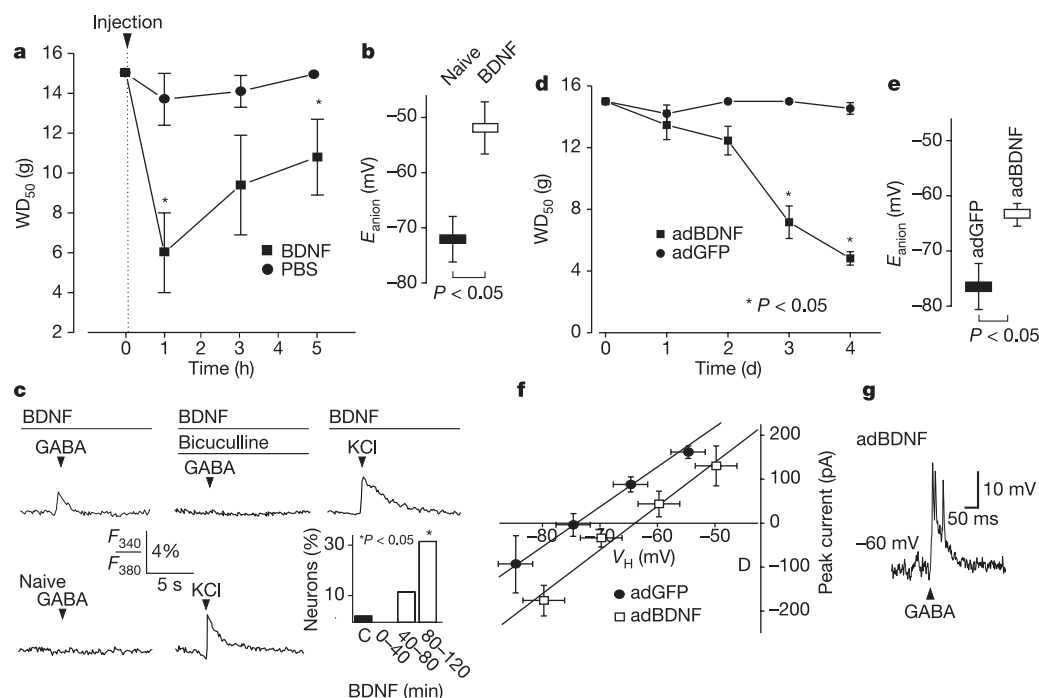


Figure 2 | Enhanced BDNF in the dorsal horn elicits nociceptive hypersensitivity and a depolarizing shift in E_{anion} in spinal LI neurons.

a, Intrathecal delivery of recombinant human BDNF (20 μ g) to the lumbar dorsal horn of intact rats led to a significant and transient decrease in the mean $\pm$ s.e.m. WD_{50} within 1 h, as compared with saline control. **b**, Significant depolarization of mean $\pm$ s.e.m. E_{anion} in LI neurons in slices treated with BDNF (50 ng ml^{-1} ; for >90 min) as compared with slices in control ACSF (Naive). **c**, Representative traces of Ca^{2+} measurements from LI neurons showing that brief GABA application in slices superfused with BDNF caused a bicuculline-sensitive increase in $[Ca^{2+}]_i$. The viability of cells not responding to GABA was confirmed by KCl-mediated responses. Bottom right inset, the proportion of LI neurons showing a GABA-mediated rise in $[Ca^{2+}]_i$ increased progressively, reaching 31% between 80 and

120 min of continuous BDNF perfusion ($\chi^2_{corrected} = 5.15$). By contrast, only 2% of cells responded with a rise in $[Ca^{2+}]_i$ over a similar time period in the absence of BDNF (C; $\chi^2_{corrected} = 6.74$). **d**, Intrathecal administration of adBDNF¹¹ triggered a delayed and progressive decrease in the mean $\pm$ s.e.m. WD_{50} that persisted as long as 4 d after injection. By contrast, administration of a control adenovirus (adGFP) elicited no decrease in WD_{50} . **e**, Significant depolarization of the mean $\pm$ s.e.m. E_{anion} in LI neurons in slices taken from the rats in **d**. **f**, Mean $\pm$ s.e.m. peak current evoked by GABA measured in LI neurons at various values of V_m in slices taken from the rats in **d**. Horizontal standard error bars represent interneuron differences. **g**, Representative trace in current-clamp recording mode showing that brief GABA application to a LI neuron in a slice taken from an adBDNF-treated rat elicited action potentials.

not vehicle control ($n = 4$), caused release of BDNF from microglia in culture ($P < 0.001$; Fig. 4e). This effect of ATP was blocked by treating the cultures with the P2X receptor blocker TNP-ATP ($n = 3$; $P < 0.05$). In addition, pretreatment of the microglia with BDNF siRNA prevented release of BDNF by ATP stimulation ($n = 3$;

$P < 0.001$). Together with our behavioural and electrophysiological results, these findings show that both the decrease in paw withdrawal threshold and the shift in E_{anion} in LI neurons caused by ATP-stimulated microglia require BDNF–TrkB signalling and that the source of BDNF is the microglia themselves.

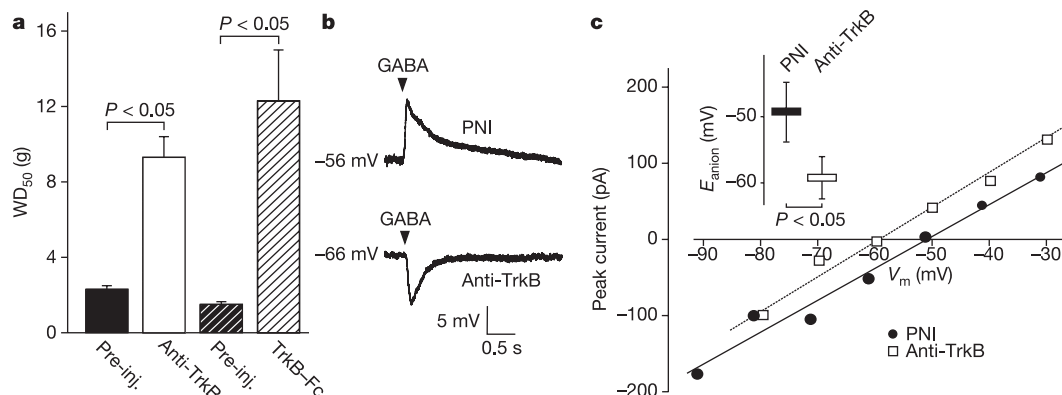


Figure 3 | Functional inhibition of BDNF–TrkB signalling reverses allodynia and the depolarizing shift in E_{anion} in spinal LI neurons in rats with PNI.

a, Intrathecal administration of either anti-TrkB or TrkB–Fc to the lumbar dorsal horn of rats that showed a robust allodynia in response to PNI caused a significant increase in the mean $\pm$ s.e.m. WD_{50} . **b**, Representative traces in current-clamp recording mode showing the postsynaptic response to GABA in LI neurons taken from PNI rats in slices treated with or without

anti-TrkB. **c**, Representative current–voltage plots in voltage-clamp recording mode of responses to brief (10-ms) local applications of GABA in two LI neurons in slices taken from PNI rats: one from a slice superfused with control ACSF, the other from a slice after anti-TrkB perfusion (1 μ g ml^{-1}). Inset, pooled data showing that anti-TrkB perfusion of slices taken from rats with PNI elicited a significant hyperpolarization of the mean $\pm$ s.e.m. E_{anion} in LI neurons.

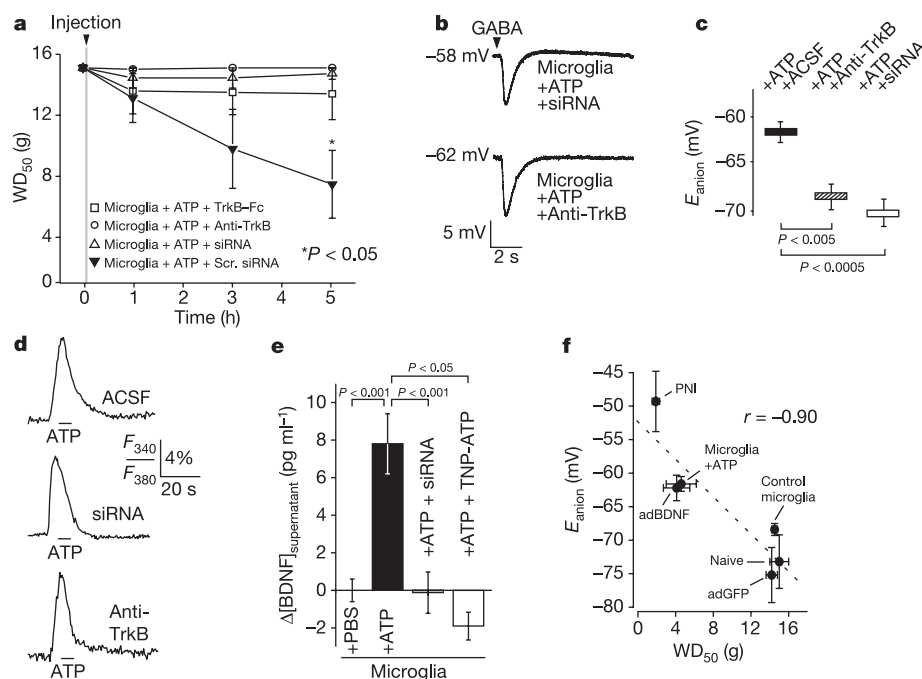


Figure 4 | Microglia-derived BDNF triggers both allodynia and the depolarizing shift in E_{anion} in LI neurons. **a**, Local spinal delivery of ATP-stimulated microglia either incubated with anti-TrkB or TrkB-Fc, or transfected with BDNF siRNA did not cause a significant change in the mean $\pm$ s.e.m. WD_{50} . By contrast, ATP-stimulated microglia transfected with a scrambled version of the interfering RNA (Scr. siRNA) caused the WD_{50} to drop significantly after 5 h. **b**, Representative traces in current-clamp recording mode showing postsynaptic responses to GABA in LI neurons from rats treated as in **a**. **c**, Pooled data showing the mean $\pm$ s.e.m. E_{anion} measured in LI neurons from rats treated as in **a**. **d**, Representative

traces of Ca^{2+} measurements from microglia showing that responses to brief applications of ATP were not affected by exposure of microglia to anti-TrkB or to BDNF siRNA. **e**, ELISA-based measurement of the mean $\pm$ s.e.m. BDNF protein in the supernatant of cultured microglia 5 h after treatment with PBS vehicle, ATP, ATP plus TNP-ATP (10 μM) or ATP after pretreatment with BDNF siRNA. **f**, Correlation plot showing the relationship between E_{anion} and WD_{50} (error bars indicate the s.e.m.). The data shown include only those where both WD_{50} and E_{anion} were recorded in the same rat.

Our present findings suggest that, after nerve injury, ATP stimulation of microglia and the subsequent BDNF–TrkB signalling are crucial elements leading to the shift in E_{anion} of LI neurons. We therefore reasoned that inhibiting microglia ATP signalling should suppress the shift in E_{anion} caused by PNI. To test this, we bath-applied TNP-ATP (which reverses nerve-injury-induced tactile allodynia by acting on P2X receptors in microglia⁴) to spinal slices taken from allodynic rats 2 weeks after PNI. In the presence of TNP-ATP, E_{anion} of LI neurons was -59.3 ± 1.8 mV ($n = 6$), which was significantly more negative than that in LI neurons from untreated slices taken from nerve-injured rats (-49.3 ± 4.5 mV; $n = 6$; $P < 0.05$). Thus, P2X receptor activation is necessary to sustain the depolarizing shift in E_{anion} in rats with PNI. Moreover, we found an inverse correlation between paw withdrawal threshold and E_{anion} in LI neurons across all experimental conditions (Fig. 4f), suggesting that E_{anion} is an essential mechanistic link between microglia and allodynia.

Because acutely disrupting the microglia–BDNF–neuron signalling pathway reversed the change in E_{anion} and allodynia, ongoing activation of this pathway must be necessary to maintain nerve-injury-induced pain. Although BDNF has been implicated in spinal nociceptive hypersensitivity, the current framework of thinking on the role of BDNF is based on the hypothesis that the source of this neurotrophin is primary afferent neurons (reviewed in ref. 15). This source of BDNF has been brought into question in neuropathic pain, however, because of evidence indicating that there is a lack of primary-afferent-evoked release of BDNF in the spinal cord after PNI¹⁶. By showing that BDNF derives from microglia, our findings change the model for understanding how BDNF participates in the nociceptive hypersensitivity that underlies tactile allodynia. Our findings provide a new perspective from which to understand the

aetiology of pain hypersensitivity and suggest that targeting microglia-derived BDNF may be a previously unsuspected avenue for treating neuropathic pain.

Our results define a biochemical pathway and a biophysical mechanism by which activated microglia affect neuronal function. Microglia become activated in many pathological processes in the central nervous system^{17–20}. Given that depolarizing anion currents can gate plasticity and excitotoxicity²¹, the microglia–neuron signalling mechanism that we have identified may be a substrate underlying these pathological disorders.

METHODS

Model of PNI and behavioural studies. PNI was induced as described^{5,22}. In brief, a polyethylene cuff (~2 mm in length) was surgically implanted around the sciatic nerve of adult male Sprague–Dawley rats. The 50% paw withdrawal threshold to mechanical stimulation was assessed as described^{5,23}. The withdrawal threshold was measured for each rat before use in slice experiments. In this model of PNI, there is microglia activation in the spinal cord ipsilateral to the nerve cuff, as indicated by a considerable increase in labelling for the microglia activation marker OX42 (by anti-CR3/CD11b, Cedarlane; diluted 1:1,000; data not shown).

Slice preparation. Parasagittal slices (300–350 μm) of spinal cord were prepared from adult male rats (aged >50 d) as described²⁴. Slices were continually superfused (2–3 ml min⁻¹) with artificial cerebrospinal fluid (ACSF) containing (in mM): 126 NaCl, 26 NaHCO₃, 10 glucose, 2.5 KCl, 2 CaCl₂, 2 MgCl₂ and 1.25 NaH₂PO₄ buffer (bubbled with 95% O₂/5% CO₂; pH $\approx$ 7.4).

Recordings. The pipettes were filled with a solution containing (in mM): 130 potassium methyl sulphate (KMeSO₄), 5 CsCl, 2 MgCl₂, 11 BAPTA, 1 CaCl₂, 4 ATP, 0.4 GTP and 10 HEPES buffer (pH $\approx$ 7.4). For perforated-patch recordings, 25 $\mu\text{g ml}^{-1}$ gramicidin D, from a stock solution of 10 mg ml⁻¹ in dimethylsulphoxide (DMSO), was added to the above solution. GABA was applied locally for 10–50 ms by pressure ejection through a micropipette. Data acquisition and analysis of postsynaptic currents (PSCs) were done as

described²⁵. Membrane potential measurements were corrected as described²⁶. Neither input resistance nor resting membrane potential of LI neurons was affected significantly by any of the drugs or protocols used in this study. All data are given as the mean $\pm$ s.e.m., except where indicated. Statistical significance was tested by using Student's *t*-test for comparison of mean values, χ^2 tests for contingency tables, and mixed-design analyses of variance (*post hoc* Tukey's HSD test) for repeated measures.

Microglia cultures. Rat primary cultures were prepared from neonatal cortex or spinal cord as indicated, under standard conditions as described^{4,27}, and maintained for 10–14 d in DMEM medium with 10% fetal bovine serum. Microglia were separated from the primary culture by gentle shaking of the flask and were replated on plastic dishes. The cells were removed from the dish surface with a cell scraper and collected in 100 μ l of PBS buffer. The density of microglia was measured by using a cell counter, and the volume of PBS was adjusted to give a final density of 1,000 cells per 10 μ l. This method produces microglia cultures of >95% purity. For ATP stimulation, the purified microglia were incubated with 50 μ M ATP for 1 h.

Intrathecal injections. At least 3 d before drug administration, rats were anaesthetized with sodium pentobarbital (65 mg per kg (body weight)), and a lumbar spinal catheter (PE-10 polyethylene tube) was inserted into the intrathecal space as described²⁸. After drug or vehicle administration, rats were killed and their vertebral column was dissected to confirm visually that placement of the catheter was correct. Drugs included BDNF (10 μ g per injection), anti-TrkB (30 μ g per injection) and TrkB-Fc (5 μ g per injection), all of which were prepared in saline plus 10% (v/v) DMSO. For virus-mediated transduction, adenoviral vectors encoding BDNF and enhanced green fluorescent protein (EGFP)¹¹ were administered once (20 μ l, at 2.0×10^{10} plaque-forming units per ml). At the doses used, none of the compounds produced motor disturbances or sedation, as assessed by grasping, righting and placing reflexes and behavioural observations²⁹. In all experiments, 30 μ l of microglia plus supernatant were injected intrathecally in normal rats, and for all drugs and microglia tested the experimenter was blind to the treatment.

RNA interference. Microglia cultures were transfected with siRNA directed against BDNF or scrambled siRNA (Dharmacon) with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. In brief, siRNA and Lipofectamine were diluted in serum-free medium, mixed and added to the microglia cultures. Transfection was allowed to occur for 5 h and the microglia were collected as described above for subsequent stimulation and intrathecal injection. The BDNF siRNA consisted of four pooled 21-nucleotide duplexes. The sequences of the four duplexes¹⁴ were TCGAAGAGCTGCTGGATGA, TATGTACACTGACCATTAA, GAGCGTGTGTGACAGTATT and GAACTACCC AATCGTATGT.

Enzyme-linked immunosorbent assay. To measure BDNF secretion, we prepared microglia under the various experimental conditions described above and incubated them at 37 °C for 6 h to model the above *in vivo* experiments. Supernatants were removed and BDNF protein concentration was measured by ELISA (Chemicon).

Calcium imaging. Spinal cord slices were prepared for Ca^{2+} imaging and tested for responses to GABA as described⁵. Primary cultures of microglia were prepared as above, transferred to standard coverslips and incubated with 2.5 μ M Fura-2-AM in HEPES-buffered saline plus 0.01% DMSO for 45 min. After fluorophore loading, changes in $[\text{Ca}^{2+}]_i$ in individual microglia were evoked with brief (~5 s) applications of ATP (10 μ M) from a micropipette. $[\text{Ca}^{2+}]_i$ was fluorometrically measured with a 40X water-immersion objective on a Zeiss Axioscope equipped with epifluorescence optics. Images were acquired by using a TILL Photonics monochromator coupled to a CCD camera, and regions of interest (for ratio analysis) were drawn on clearly distinct cell bodies.

Received 13 June; accepted 8 September 2005.

1. Woolf, C. J. & Salter, M. W. Neuronal plasticity: increasing the gain in pain. *Science* **288**, 1765–1769 (2000).
2. Ji, R. R. & Strichartz, G. Cell signalling and the genesis of neuropathic pain. *Sci. STKE* **2004**, reE14 (2004).
3. Lewin, G. R., Lu, Y. & Park, T. J. A plethora of painful molecules. *Curr. Opin. Neurobiol.* **14**, 443–449 (2004).
4. Tsuda, M. *et al.* P2X4 receptors induced in spinal microglia gate tactile allodynia after nerve injury. *Nature* **424**, 778–783 (2003).
5. Coull, J. A. *et al.* Trans-synaptic shift in anion gradient in spinal lamina I neurons as a mechanism of neuropathic pain. *Nature* **424**, 938–942 (2003).
6. Nakajima, K., Tohyama, Y., Kohsaka, S. & Kurihara, T. Ceramide activates microglia to enhance the production/secretion of brain-derived neurotrophic factor (BDNF) without induction of deleterious factors *in vitro*. *J. Neurochem.* **80**, 697–705 (2002).

7. Mannion, R. J. *et al.* Neurotrophins: peripherally and centrally acting modulators of tactile stimulus-induced inflammatory pain hypersensitivity. *Proc. Natl Acad. Sci. USA* **96**, 9385–9390 (1999).
8. Heppenstall, P. A. & Lewin, G. R. BDNF but not NT-4 is required for normal flexion reflex plasticity and function. *Proc. Natl Acad. Sci. USA* **98**, 8107–8112 (2001).
9. Thompson, S. W., Bennett, D. L., Kerr, B. J., Bradbury, E. J. & McMahon, S. B. Brain-derived neurotrophic factor is an endogenous modulator of nociceptive responses in the spinal cord. *Proc. Natl Acad. Sci. USA* **96**, 7714–7718 (1999).
10. Rivera, C. *et al.* BDNF-induced TrkB activation down-regulates the K^+ – Cl^- cotransporter KCC2 and impairs neuronal Cl^- extrusion. *J. Cell Biol.* **159**, 747–752 (2002).
11. Gravel, C., Gotz, R., Lorrain, A. & Sendtner, M. Adenoviral gene transfer of ciliary neurotrophic factor and brain-derived neurotrophic factor leads to long-term survival of axotomized motor neurons. *Nature Med.* **3**, 765–770 (1997).
12. Balkowiec, A. & Katz, D. M. Activity-dependent release of endogenous brain-derived neurotrophic factor from primary sensory neurons detected by ELISA *in situ*. *J. Neurosci.* **20**, 7417–7423 (2000).
13. Jiang, B., Akaneya, Y., Hata, Y. & Tsumoto, T. Long-term depression is not induced by low-frequency stimulation in rat visual cortex *in vivo*: a possible preventing role of endogenous brain-derived neurotrophic factor. *J. Neurosci.* **23**, 3761–3770 (2003).
14. Baker-Herman, T. L. *et al.* BDNF is necessary and sufficient for spinal respiratory plasticity following intermittent hypoxia. *Nature Neurosci.* **7**, 48–55 (2004).
15. Malcangio, M. & Lessmann, V. A common thread for pain and memory synapses? Brain-derived neurotrophic factor and trkB receptors. *Trends Pharmacol. Sci.* **24**, 116–121 (2003).
16. Lever, I., Cunningham, J., Grist, J., Yip, P. K. & Malcangio, M. Release of BDNF and GABA in the dorsal horn of neuropathic rats. *Eur. J. Neurosci.* **18**, 1169–1174 (2003).
17. Nguyen, M. D., Julien, J. P. & Rivest, S. Innate immunity: the missing link in neuroprotection and neurodegeneration? *Nature Rev. Neurosci.* **3**, 216–227 (2002).
18. Streit, W. J., Walter, S. A. & Pennell, N. A. Reactive microgliosis. *Prog. Neurobiol.* **57**, 563–581 (1999).
19. Kreutzberg, G. W. Microglia: a sensor for pathological events in the CNS. *Trends Neurosci.* **19**, 312–318 (1996).
20. Gonzalez-Scarano, F. & Baltuch, G. Microglia as mediators of inflammatory and degenerative diseases. *Annu. Rev. Neurosci.* **22**, 219–240 (1999).
21. Staley, K. J., Soldo, B. L. & Proctor, W. R. Ionic mechanisms of neuronal excitation by inhibitory GABA_A receptors. *Science* **269**, 977–981 (1995).
22. Mosconi, T. & Kruger, L. Fixed-diameter polyethylene cuffs applied to the rat sciatic nerve induce a painful neuropathy: ultrastructural morphometric analysis of axonal alterations. *Pain* **64**, 37–57 (1996).
23. Chaplan, S. R., Bach, F. W., Pogrel, J. W., Chung, J. M. & Yaksh, T. L. Quantitative assessment of tactile allodynia in the rat paw. *J. Neurosci. Methods* **53**, 55–63 (1994).
24. Chery, N., Yu, X. H. & De Koninck, Y. Visualization of lamina I of the dorsal horn in live adult rat spinal cord slices. *J. Neurosci. Methods* **96**, 133–142 (2000).
25. Keller, A. F., Coull, J. A., Chery, N., Poiseau, P. & De Koninck, Y. Region-specific developmental specialization of GABA-glycine cosynapses in laminae I–II of the rat spinal dorsal horn. *J. Neurosci.* **21**, 7871–7880 (2001).
26. Tyzio, R. *et al.* Membrane potential of CA3 hippocampal pyramidal cells during postnatal development. *J. Neurophysiol.* **90**, 2964–2972 (2003).
27. Nakajima, K. *et al.* Identification of elastase as a secretory protease from cultured rat microglia. *J. Neurochem.* **58**, 1401–1408 (1992).
28. Yaksh, T. L., Jessell, T. M., Gamse, R., Mudge, A. W. & Leeman, S. E. Intrathecal morphine inhibits substance P release from mammalian spinal cord *in vivo*. *Nature* **286**, 155–157 (1980).
- 29.Coderre, T. J. & Van Empel, I. The utility of excitatory amino acid (EAA) antagonists as analgesic agents I. Comparison of the antinociceptive activity of various classes of EAA antagonists in mechanical, thermal and chemical nociceptive tests. *Pain* **59**, 345–352 (1994).

Acknowledgements We thank F. Nault for technical assistance; J. Zhang for conducting OX42 immunostaining; and P. De Koninck, M. Cordero-Erasquin, F. Keller and C. Labrakakis for discussions on the manuscript. This work was supported by Canadian Institutes of Health Research (CIHR) grants (to Y.D.K. and to M.W.S.), Brain Repair Program of Neuroscience Canada and its partner the Ontario Neurotrauma Foundation through the BRP. J.A.M.C. was the recipient of a doctoral award from the CIHR; Y.D.K. is a Senior Scholar of the Fonds de la recherche en santé du Québec (FRSQ); S.B. was the recipient of a fellowship from a Strategic Training Program Grant from the CIHR.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.D.K. (Yves.DeKoninck@curlg.ulaval.ca) or M.W.S. (mike.salter@utoronto.ca).

LETTERS

Heat activation of TRPM5 underlies thermal sensitivity of sweet taste

Karel Talavera¹, Keiko Yasumatsu², Thomas Voets¹, Guy Droogmans¹, Noriatsu Shigemura², Yuzo Ninomiya², Robert F. Margolskee³ & Bernd Nilius¹

TRPM5, a cation channel of the TRP superfamily, is highly expressed in taste buds of the tongue, where it has a key role in the perception of sweet, umami and bitter tastes^{1,2}. Activation of TRPM5 occurs downstream of the activation of G-protein-coupled taste receptors and is proposed to generate a depolarizing potential in the taste receptor cells². Factors that modulate TRPM5 activity are therefore expected to influence taste. Here we show that TRPM5 is a highly temperature-sensitive, heat-activated channel: inward TRPM5 currents increase steeply at temperatures between 15 and 35 °C. TRPM4, a close homologue of TRPM5, shows similar temperature sensitivity. Heat activation is due to a temperature-dependent shift of the activation curve, in analogy to other thermosensitive TRP channels³. Moreover, we show that increasing temperature between 15 and 35 °C markedly enhances the gustatory nerve response to sweet compounds in wild-type but not in *Trpm5* knockout mice. The strong temperature sensitivity of TRPM5 may underlie known effects of temperature on perceived taste in humans^{4–6}, including enhanced sweetness perception at high temperatures and ‘thermal taste’, the phenomenon whereby heating or cooling of the tongue evoke sensations of taste in the absence of tastants⁷.

Temperature has a strong influence on how we taste. For example, the perceived sweetness of diluted sugar solutions increases strongly with temperature^{4,5}. In addition, cooling or heating of the tongue by itself is sufficient to cause sensations of taste in ~50% of humans⁷. At present, the mechanisms underlying the thermal effects on taste are fully unknown. Given that several members of the TRP superfamily function as thermosensors^{8–10}, we tested whether TRPM5 is a temperature-sensitive component in the taste pathway.

TRPM5 was transiently expressed in HEK-293 cells and studied with the whole-cell patch-clamp technique. As previous studies have shown that TRPM5 functions as an intracellular Ca^{2+} -activated, voltage-dependent channel, we included 500 nM Ca^{2+} in the intracellular solution and measured currents during voltage steps to potentials ranging from –75 to +175 mV from a holding potential of +25 mV. Under these conditions, increasing the bath temperature from 14 to 38 °C led to a marked increase in current amplitude and the current relaxation rate (Fig. 1a), showing that TRPM5 is a heat-activated channel. The 10-degree temperature coefficient (Q_{10}) for current activation at –75 mV amounted to 10.3 ± 2.0 ($n = 6$) between 15 and 25 °C, which is in the same range as the Q_{10} values of other thermosensitive TRPs (thermoTRPs)^{8,9}. The effect of temperature on current amplitude was strongly voltage dependent: inward current at –75 mV was undetectable at 14 °C, whereas at the same temperature robust outward currents were measured at +75 and +150 mV (Fig. 1b, c). Heat activation results from a strong shift of the voltage-dependent activation curve to negative potentials,

with a change in the voltage for half-maximal activation (V_{act}) of -7.0 ± 0.6 mV per 1 °C increase in temperature (Fig. 1d, e). The maximal whole-cell conductance increased with temperature with an estimated Q_{10} of 1.17 ± 0.01 , indicating that the temperature

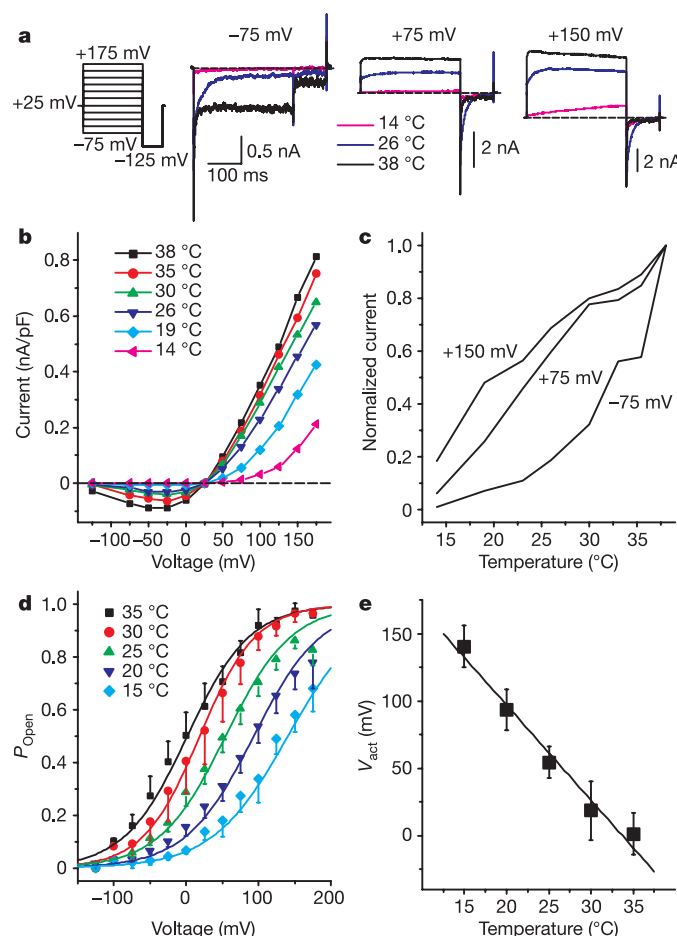


Figure 1 | Heat activation of TRPM5. **a**, Voltage protocol and current traces evoked at different voltages at 14, 26 and 38 °C. **b**, Current–voltage relationships at different temperatures. **c**, Temperature dependence of the current amplitudes at different potentials, normalized to the values at 38 °C. **d**, Average activation curves at different temperatures ($n = 4–7$). Lines are Boltzmann functions calculated with average values of the voltage for half-maximal activation (V_{act}) and slope factor (s_{act}) obtained for each temperature. **e**, Temperature dependence of V_{act} . Data are the mean $\pm$ s.e.m.

¹Laboratorium voor Fysiologie, Campus Gasthuisberg, KU Leuven, B-3000 Leuven, Belgium. ²Section of Oral Neuroscience, Graduate School of Dental Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan. ³Department of Neuroscience, Mount Sinai School of Medicine, 1425 Madison Avenue, New York, New York 10029, USA.

dependence of the single-channel conductance resembles to that of aqueous diffusion¹¹.

Given that TRPM5 is modulated by intracellular Ca^{2+} (refs 12–14), we tested whether the effects of temperature on channel activation were Ca^{2+} dependent. Heat activation was abolished when 10 mM EGTA ($[\text{Ca}^{2+}]_i < 1 \text{ nM}$) was included in the patch pipette (Fig. 2a), indicating that cytosolic Ca^{2+} is absolutely required for channel activation. In contrast, heat activation was fully preserved when intracellular Ca^{2+} was set at $100 \mu\text{M}$, a concentration that is around 100-fold larger than needed for maximal Ca^{2+} -dependent activation of TRPM5 under our experimental conditions¹⁴ (Fig. 2b). Heating still induced a shift of the activation curve to negative potentials ($-7.9 \pm 0.6 \text{ mV}^\circ\text{C}^{-1}$; Fig. 2d, f) and an increase in the rate of current relaxation at every membrane potential (Fig. 2e). Together, these data indicate that heat activation of TRPM5 requires cytosolic Ca^{2+} , but is not due to temperature-dependent effects on the Ca^{2+} -dependent activation step. Heat activation of TRPM5 was also observed in cell-free inside-out patches, indicating that temperature modulation occurs in a membrane-delimited manner (see Supplementary Fig. S1).

TRPM4, which is the closest homologue of TRPM5, showing more

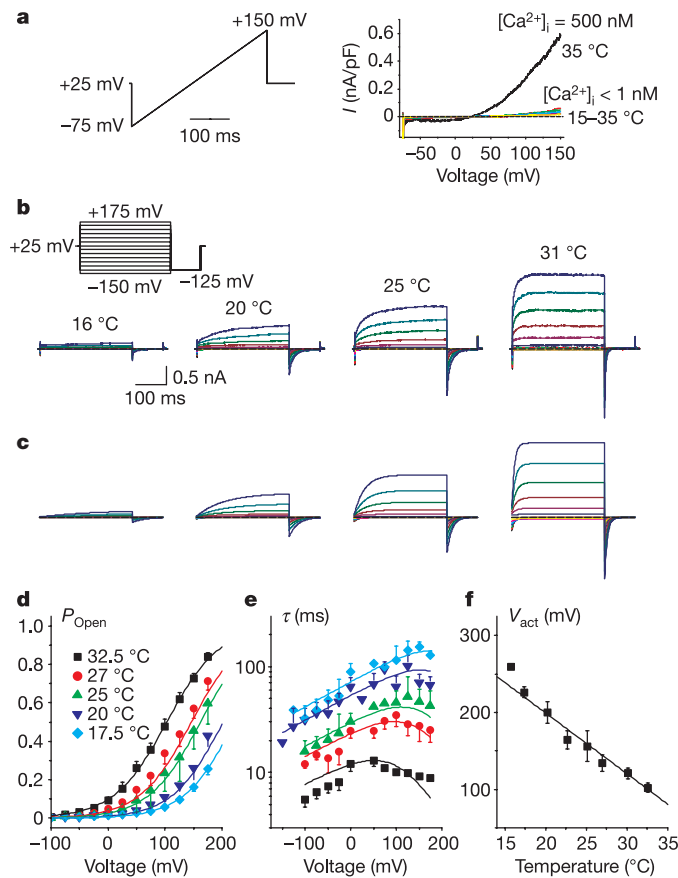
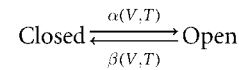


Figure 2 | Temperature dependence of TRPM5 in a high intracellular Ca^{2+} concentration. **a**, TRPM5 currents elicited by a voltage ramp (left) at an intracellular Ca^{2+} concentration buffered to 500 nM (35 °C, black) or below 1 nM (10 mM EGTA, colours) and at temperatures from 15 to 35 °C. **b**, Current traces elicited at different temperatures in response to the indicated voltage protocol at an intracellular Ca^{2+} concentration of $100 \mu\text{M}$. **c**, Simulated current traces at different temperatures as predicted by the model for TRPM5 (see Supplementary Information). **d**, **e**, Average activation curves and voltage dependence of the time constant of current relaxation at different temperatures ($n = 3$ –5). **f**, Temperature dependence of V_{act} . Continuous lines represent the fit of the two-state gating model in **e** and **f**, and the activation curves predicted by the model in **d**. Data are the mean $\pm$ s.e.m.

than 50% sequence similarity at the amino acid level and sharing the same functional hallmarks such as voltage dependence and activation by Ca^{2+} , may therefore show analogous temperature dependence. We measured TRPM4 currents in inside-out patches and at a saturating intracellular Ca^{2+} concentration of 1 mM to minimize current desensitization¹⁵. Under these conditions, heating strongly enhanced TRPM4 currents (see Supplementary Fig. S2a). Current amplitude at +25 mV showed a Q_{10} of 8.5 ± 0.6 between 15 and 25 °C. Heating shifted V_{act} towards physiological voltages ($-5.6 \pm 0.5 \text{ mV}^\circ\text{C}^{-1}$; Supplementary Fig. S2c, e) and increased the rate of current relaxation at every membrane potential (Supplementary Fig. S2d). The maximal whole-cell conductance increased with temperature with a Q_{10} of 1.32 ± 0.08 . Temperature had little effect on the Ca^{2+} dependence of channel activation (Supplementary Fig. S2f): the Ca^{2+} concentration for half-maximal activation decreased only slightly on heating from 15 °C to 37 °C, with an estimated Q_{10} of 1.36 ± 0.02 (Supplementary Fig. S2f, inset). Thus, as for TRPM5, the strong temperature dependence of TRPM4 is not due to a modulation of the Ca^{2+} dependency of the channel.

We have previously shown that thermal activation of the heat sensor TRPV1 and the cold sensor TRPM8 can be described by a simple two-state model for a voltage-gated channel³:



where α and β are the voltage- and temperature-dependent opening and closing rates, respectively. A fit of the experimental data (Fig. 2d–f and Supplementary Fig. S2c–e) allowed us to determine the enthalpies and entropies associated with the opening (ΔH_α and ΔS_α) and closing (ΔH_β and ΔS_β) of TRPM4 and TRPM5, along with the apparent gating valence (z) and the electrical coupling factor (δ ; Fig. 3a–d). Simulations using these parameters yielded a close approximation of the experimental results for TRPM5 and TRPM4 (Fig. 2c, d, and Supplementary Fig. S2b, c). Opening of TRPM4 and TRPM5 is associated with a significant increase in both enthalpy and entropy (large positive values of $\Delta H^\ddagger$ and $\Delta S^\ddagger$; Fig. 3a, b), and thus obeys the same principle as the heat activation of TRPV1 (ref. 3). In contrast, opening of the cold-activated TRPM8 is associated with a large decrease in both enthalpy and entropy. The two-state kinetic model predicts that the midpoint of the activation curves is related to temperature as follows: $V_{\text{act}} = (\Delta H^\ddagger - T\Delta S^\ddagger)/zF$ (see Supplementary Information). TRPM4 and TRPM5, like TRPM8 and TRPV1 (ref. 3), combine a large absolute value of $\Delta S^\ddagger$ with a low gating valence z (Fig. 3b, c), which explains the large shifts in V_{act} on changing temperature (Fig. 3e). ‘Classical’ voltage-gated channels have a much larger z , and, accordingly, show a much less pronounced temperature sensitivity^{16–18}. Notably, the structural determinants for the strong temperature dependence are conserved over TRPM and TRPV subfamilies, in contrast to those that determine whether these channels are heat or cold activated¹⁹.

Measurable heat activation of inward TRPM5 currents at physiological potentials occurs above $\sim 15^\circ\text{C}$ in a temperature range between that of significant activation of TRPM8 and that of TRPV1 (Fig. 3f). Notably, this range spans across room temperatures, in the apparent gap of temperature activation of thermoTRPs known so far². Two other striking features set TRPM4 and TRPM5 apart from the other thermosensitive TRPs. First, TRPM4 and TRPM5 are the only Ca^{2+} -impermeable thermoTRPs described so far. As a consequence, heat activation of these channels leads to membrane depolarization but does not directly cause an increase in intracellular Ca^{2+} . Second, activation of TRPM4 and TRPM5 is abolished when intracellular Ca^{2+} is buffered to low levels. These channels thus have the potential to act as coincidence detectors of warm temperatures and stimuli that increase intracellular Ca^{2+} .

It is well known that temperature modulates human perception of

different taste modalities, including those involving TRPM5. The clearest example is the strong enhancement of perceived sweetness with increasing temperature^{4,5}. To determine whether the temperature sensitivity of TRPM5 modulates taste perception *in vivo*, we carried out electrical recordings in chorda tympani nerves from wild-type and *Trpm5* knockout mice. These nerves innervate the anterior taste field of the tongue, which is preferentially responsive to sweet over bitter and umami tastes. Chorda tympani nerve stimulation in wild-type mice by two sugars (sucrose and glucose) and two non-caloric sweeteners (saccharin and SC45647) induced dose-dependent responses that increased ~3–10-fold on increasing the temperature from 15 to 35 °C (Fig. 4a–d). Similar temperature-dependent increments were observed with two other sugars (maltose, fructose) and the sweet amino acid glycine, but not with umami (monosodium glutamate), sour (HCl), salty (NaCl), or bitter (quinine hydrochloride) stimuli (Fig. 4e).

In comparison to wild type, chorda tympani responses to sweet compounds were greatly diminished (all sugars, saccharin and glycine) or absent (SC45647) in *Trpm5* knockout mice (Fig. 4f). In marked contrast to wild type, the residual responses of the *Trpm5* knockout mice to all of the sweet compounds tested did not increase significantly on changing the temperature from 15 to 35 °C. Again,

there were no temperature-dependent increments in the chorda tympani responses of the *Trpm5* knockout mice to monosodium glutamate, HCl, NaCl or quinine hydrochloride. Notably, integrated data from chorda tympani and glossopharyngeal nerves, together with results from behavioural assays, indicate that responses to sweet, bitter and umami tastes are severely diminished but not completely impaired in our *Trpm5* knockout model (S. Damak, M. Rong, K.Y., Z. Kokrashvili, C. Perez, N.S., R. Yoshida, B. Mosinger, J. Glendinning, Y.N. & R. Margolskee, unpublished data). Therefore, our results differ quantitatively from a previous study², which showed a complete loss of sweet, bitter and umami responses in a *Trpm5* knockout model. This discrepancy could be due to differences between the knockout targeting constructs (S. Damak, M. Rong, K.Y., Z. Kokrashvili, C. Perez, N.S., R. Yoshida, B. Mosinger, J. Glendinning, Y.N. & R. Margolskee, unpublished data).

In conclusion, we have identified TRPM4 and TRPM5 as two temperature-sensitive cation channels that are activated by heating in

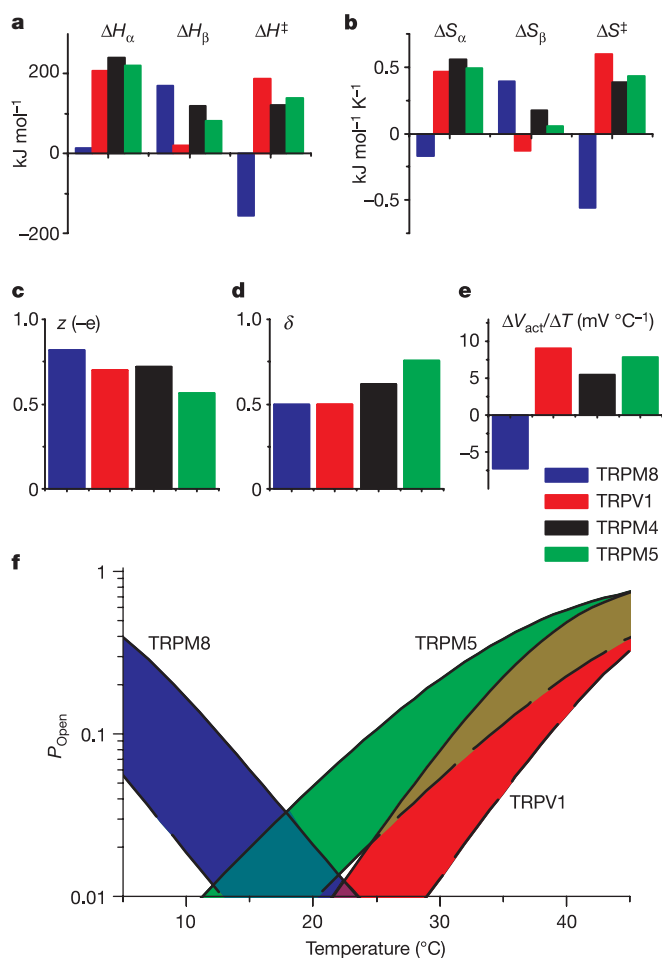


Figure 3 | Comparison of the properties of TRPM8, TRPV1, TRPM4 and TRPM5. **a, b**, Enthalpies and entropies associated with the opening (ΔH_α and ΔS_α) and closing (ΔH_β and ΔS_β) transitions and differences in enthalpy and entropy between the open and the closed states ($\Delta H^\ddagger$ and $\Delta S^\ddagger$). **c**, Effective valence of the gating charge (z). **d**, Coupling electric factor (δ). **e**, Shift of V_{act} per degree of temperature ($\Delta V_{\text{act}}/\Delta T$). **f**, Activity of TRPM8, TRPV1 and TRPM5 (Ca^{2+} 500 nM) as a function of temperature. Coloured bands correspond to open probabilities in the range –100 mV (broken lines) to –30 mV (unbroken lines).

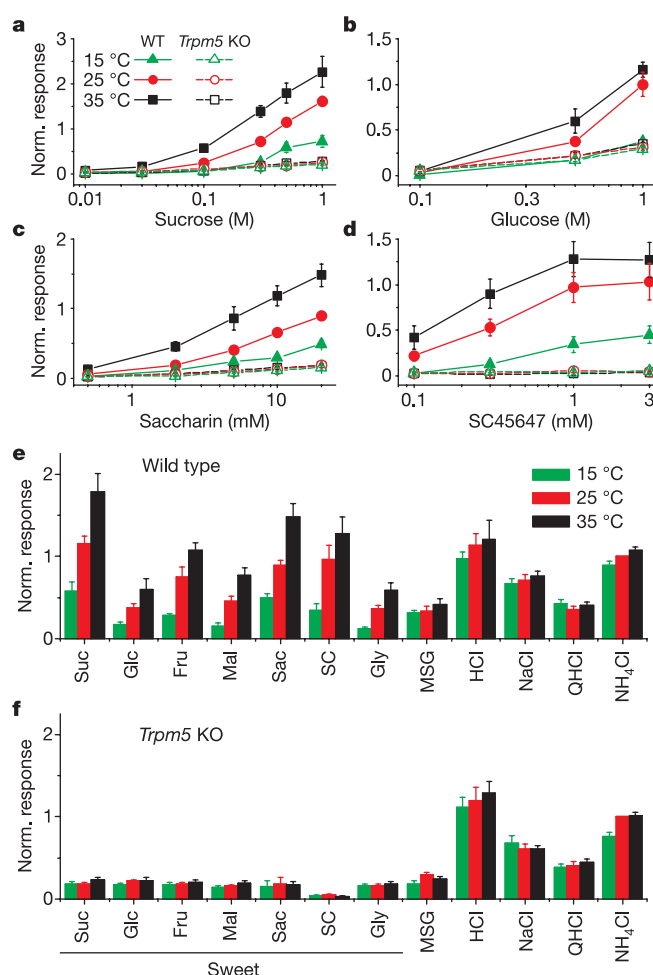


Figure 4 | Temperature dependence of chorda tympani nerve responses of wild-type and *Trpm5* knockout mice. **a–d**, Dose-dependent effects of the sweet compounds sucrose, glucose, saccharin and SC45647 on whole-nerve recordings from chorda tympani nerve of wild-type and *Trpm5* knockout mice at 15, 25 and 35 °C ($n = 6–8$). **e, f**, Whole-nerve responses from chorda tympani nerve of wild-type and *Trpm5* knockout mice stimulated by sweet, umami, sour, salty and bitter compounds at 15, 25 and 35 °C ($n = 6–8$). Sweet compounds: sucrose, 0.5 M (Suc); glucose, 0.5 M (Glc); fructose, 0.5 M (Fru); maltose, 0.5 M (Mal); saccharin, 20 mM (Sac); SC45647, 1 mM (SC); glycine, 0.3 M (Gly). Umami compound: monosodium glutamate, 0.1 M (MSG). Sour compound: HCl, 10 mM. Salty compound: NaCl, 0.1 M. Bitter compound: quinine hydrochloride, 20 mM (QHCl). Nerve responses were normalized to the response to 0.1 M NH_4Cl at 25 °C. Data are the mean $\pm$ s.e.m.

the temperature range between 15 and 35 °C. TRPM5 is involved in the transduction of taste, probably by providing a depolarizing current that is activated downstream of taste receptor activation^{1,2}. Accordingly, we found that TRPM5-dependent taste responses in mouse gustatory nerves are strongly enhanced by heating. The thermal sensitivity of TRPM5 may also explain several known effects of temperature on taste perception in humans, such as the stronger perceived sweetness of sugar solutions at warmer temperatures^{4–6}. Moreover, direct heat activation of TRPM5 could lead to activation of taste receptor cells in the absence of tastants, providing a straightforward explanation for the phenomenon of thermal taste⁷.

METHODS

Patch-clamp experiments. *Trpm4b* and *Trpm5* were cloned in the pCAGGS–IRES–GFP vector and transiently transfected into HEK-293 cells by using TransIT-293 reagents (Mirus). Cell culture and patch-clamp recordings were done as described²⁰. To minimize current rundown, TRPM4 and TRPM5 currents were recorded under inside-out and whole-cell conditions, respectively¹⁴. Holding potentials differed from zero and were such that no steady ionic flows occurred between test pulses (TRPM4, –50 mV, channels are closed; TRPM5, +25 mV, which equals the reversal potential). This allowed us to monitor leak currents and to avoid ionic depletion and accumulation in the intracellular milieu. The extracellular solution contained (in mM): 150 NaCl, 5 CaCl₂, 1 MgCl₂ and 10 HEPES (titrated to pH 7.4 with NaOH). For experiments with TRPM4, the intracellular solutions contained 150 mM NaCl and 10 mM HEPES (titrated to pH 7.2 with NaOH). For TRPM5, 50 mM NaCl was substituted with 100 mM NMDG⁺. Intracellular Ca²⁺ was buffered by adding 2 mM EGTA (for 500 nM free Ca²⁺) or 2 mM HEDTA (for 10–30 μM free Ca²⁺) and total Ca²⁺ concentrations were calculated with the software Cabuf (G. Droogmans; <ftp://ftp.cc.kuleuven.ac.be/pub/droogmans/Cabuf.zip>). The temperature of the solutions perfusing the cells was controlled as described³.

The patch-clamp data were analysed with WinASCD (G. Droogmans; <ftp://ftp.cc.kuleuven.ac.be/pub/droogmans/winascd.zip>) and Origin 7.0 (OriginLab Corporation). Whenever possible, linear background components of current traces were digitally subtracted before data analysis. Activation curves were determined from the amplitudes of tail currents elicited by sudden repolarization from different test potentials. These amplitudes were plotted as a function of the test potential and normalized to the maximal current value (*I*_{max}) obtained from the fit of the data with a Boltzmann function of the form:

$$I_{\text{Tail}}(V) = \frac{I_{\text{max}}}{1 + \exp(-(V - V_{\text{act}})/s_{\text{act}})}$$

where *V*_{act} is the potential of half-maximal activation and *s*_{act} is the slope factor. Time constants of current relaxation (*τ*) were obtained from the fit of current traces with a single exponential function.

TRPM5 knockout mice. Generation and characterization of *Trpm5* knockout mice have been described (ref. 21; and S. Damak, M. Rong, K.Y., Z. Kokrashvili, C. Perez, N.S., R. Yoshida, B. Mosinger, J. Glendinning, Y.N. & R. Margolskee, unpublished data). The *Trpm5* gene in the knockout mice has a deletion that removes 2.4-kb of the 5′-flanking region of the gene comprising the promoter and exons 1–4, including the translation start site in exon 2. By immunohistochemistry it has been shown that taste buds in these mice lack TRPM5 protein but retain the normal complement of taste receptor cells (S. Damak, M. Rong, K.Y., Z. Kokrashvili, C. Perez, N.S., R. Yoshida, B. Mosinger, J. Glendinning, Y.N. & R. Margolskee, unpublished data). The *Trpm5* knockout mice were maintained in the C57BL/6J background.

Nerve recordings. Whole-nerve responses to lingual application of tastants were recorded from the chorda tympani nerve as described²². Reservoirs of tastant solutions were adjusted to temperatures of 15, 25 or 35 °C. Before each stimulation, the tongue was pre-adapted for at least 1 min with distilled water at the temperature to be tested. The temperature of test solutions was monitored with a thermocouple placed on the tongue. Tastants were applied for 30 s. Integrated whole-nerve response magnitudes (time constant 1 s) were measured 5, 10, 15, 20 and 25 s after stimulus onset, averaged and normalized to responses to NH₄Cl at 25 °C. Significant effect of temperature on the responses to each

compound was tested with analysis of variance (ANOVA). All data are presented as the mean ± s.e.m.

Received 29 July; accepted 15 September 2005.

- Perez, C. A. *et al.* A transient receptor potential channel expressed in taste receptor cells. *Nature Neurosci.* **5**, 1169–1176 (2002).
- Zhang, Y. *et al.* Coding of sweet, bitter, and umami tastes: different receptor cells sharing similar signalling pathways. *Cell* **112**, 293–301 (2003).
- Voets, T. *et al.* The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels. *Nature* **430**, 748–754 (2004).
- Bartoshuk, L. M., Rennert, K., Rodin, J. & Stevens, J. C. Effects of temperature on the perceived sweetness of sucrose. *Physiol. Behav.* **28**, 905–910 (1982).
- Green, B. G. & Frankmann, S. P. The effect of cooling on the perception of carbohydrate and intensive sweeteners. *Physiol. Behav.* **43**, 515–519 (1988).
- Green, B. G. & Frankmann, S. P. The effect of cooling the tongue on the perceived intensity of taste. *Chem. Senses* **12**, 609–619 (1987).
- Cruz, A. & Green, B. G. Thermal stimulation of taste. *Nature* **403**, 889–892 (2000).
- Clapham, D. E. TRP channels as cellular sensors. *Nature* **426**, 517–524 (2003).
- Patapoutian, A., Peier, A. M., Story, G. M. & Viswanath, V. ThermoTRP channels and beyond: mechanisms of temperature sensation. *Nature Rev. Neurosci.* **4**, 529–539 (2003).
- Patapoutian, A. TRP channels and thermosensation. *Chem. Senses* **30** (suppl. 1), i193–i194 (2005).
- Hille, B. in *Ion Channels of Excitable Membranes* (ed. Hille, B.) Ch. 11, 366 (Sinauer Associates, Sunderland, 2001).
- Liu, D. & Liman, E. R. Intracellular Ca²⁺ and the phospholipid PIP₂ regulate the taste transduction ion channel TRPM5. *Proc. Natl Acad. Sci. USA* **100**, 15160–15165 (2003).
- Prawitt, D. *et al.* TRPM5 is a transient Ca²⁺-activated cation channel responding to rapid changes in [Ca²⁺]. *Proc. Natl Acad. Sci. USA* **100**, 15166–15171 (2003).
- Ullrich, N. D. *et al.* Comparison of functional properties of the Ca²⁺-activated cation channels TRPM4 and TRPM5 from mice. *Cell Calcium* **37**, 267–278 (2005).
- Nilius, B. *et al.* Regulation of the Ca²⁺ sensitivity of the nonselective cation channel TRPM4. *J. Biol. Chem.* **280**, 6423–6433 (2005).
- van Lunteren, E., Elmslie, K. S. & Jones, S. W. Effects of temperature on calcium current of bullfrog sympathetic neurons. *J. Physiol. Lond.* **466**, 81–93 (1993).
- Tiwari, J. K. & Sikdar, S. K. Temperature-dependent conformational changes in a voltage-gated potassium channel. *Eur. Biophys. J.* **28**, 338–345 (1999).
- Correa, A. M., Bezanilla, F. & Latorre, R. Gating kinetics of batrachotoxin-modified Na⁺ channels in the squid giant axon. Voltage and temperature effects. *Biophys. J.* **61**, 1332–1352 (1992).
- Voets, T., Talavera, K., Owsianik, G. & Nilius, B. Sensing with TRP channels. *Nature Chem. Biol.* **1**, 85–92 (2005).
- Talavera, K., Janssens, A., Klugbauer, N., Droogmans, G. & Nilius, B. Extracellular Ca²⁺ modulates the effects of protons on gating and conduction properties of the T-type Ca²⁺ channel α_{1G} (Cav3.1). *J. Gen. Physiol.* **121**, 511–528 (2003).
- Rong, M. *et al.* Signal transduction of umami taste: insights from knockout mice. *Chem. Senses* **30**, i33–i34 (2005).
- Kawai, K., Sugimoto, K., Nakashima, K., Miura, H. & Ninomiya, Y. Leptin as a modulator of sweet taste sensitivities in mice. *Proc. Natl Acad. Sci. USA* **97**, 11044–11049 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Flockerzi, R. Vennekens and U. Wissenbach for the TRPM4 clone; N. Ullrich, G. Owsianik and F. Mahieu for discussions; and R. Yoshida, A. Janssens and J. Prenen for technical assistance. This work was supported by grants from the Human Frontiers Science Program, the Belgian Federal Government (Interuniversity Poles of Attraction Program, Prime Ministers Office), the Flemish Government and the Onderzoeksrada KU Leuven, and the NIH (to R.F.M.) and by a Grant-in-Aid for scientific research from the Japan Society for the Promotion of Science (to Y.N. and to N.S.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to K.T. (Karel.talavera@med.kuleuven.be).

LETTERS

Mapping stem cell activities in the feather follicle

Zhicao Yue^{1,†}, Ting-Xin Jiang¹, Randall Bruce Widelitz¹ & Cheng-Ming Chuong¹

It is important to know how different organs ‘manage’ their stem cells. Both hair and feather follicles show robust regenerative powers that episodically renew the epithelial organ. However, the evolution of feathers (from reptiles to birds) and hairs (from reptiles to mammals) are independent events and their follicular structures result from convergent evolution. Because feathers do not have the anatomical equivalent of a hair follicle bulge, we are interested in determining where their stem cells are localized. By applying long-term label retention¹, transplantation² and DiI tracing to map stem cell activities, here we show that feather follicles contain slow-cycling long-term label-retaining cells (LRCs), transient amplifying cells and differentiating keratinocytes. Each population, located in anatomically distinct regions, undergoes dynamic homeostasis during the feather cycle. In the growing follicle, LRCs are enriched in a ‘collar bulge’ niche. In the moulting follicle, LRCs shift to populate a papillar ectoderm niche near the dermal papilla. On transplantation, LRCs show multipotentiality. In a three-dimensional view, LRCs are configured as a ring that is horizontally placed in radially symmetric feathers but tilted in bilaterally symmetric feathers. The changing topology of stem cell activities may contribute to the construction of complex feather forms.

Feathers are of great interest because their robust regenerative ability can produce different feather morphologies from common feather precursor cells. In the past decade, much was learned about hair stem cells^{1–3}. In comparison, little is known about feather stem cells. On the basis of surgery and dermal papilla transplantation, it was proposed^{4,5} that cells in the follicle collar⁶ (corresponding to the hair matrix; Fig. 1a and Supplementary Fig. 1a) have the enduring capacity to produce subsequent feather generations, but it was not known which region was essential. Papillar ectoderm, a population of keratinocytes tightly associated with the dermal papilla, was considered to have regenerative capacity; but after stripping off the papillar ectoderm, feathers could still regenerate^{7,8}. How the epidermal collar generates the feather is another long-standing unresolved issue. One view is that the collar is an anlage of a pre-patterned feather⁸. Another is that the collar cells sustain proliferation potential to generate epithelial cells that are then specified⁴.

Here we apply modern stem cell biology concepts and current technologies to revisit these fundamental biological issues. Three criteria were used to map stem and progenitor cell activities in the feather follicles during the growth and moulting phases: long-term label retention for slow-cycling cells^{9,10}, grafting for multipotentiality¹, and DiI tracing for epidermal cell migration¹¹. More than 60 yr ago, a region was identified in the collar as a “columnar epithelium of collar with elongated nuclei”, but the function of this region was not known⁴ (Supplementary Fig. 1b). Here we show that this region, which we name the ‘collar bulge’, is a niche enriched with LRCs that show stem cell characteristics. We further analyse the dynamics of LRCs and stem or progenitor cell activities during physiological feather cycling.

We began by identifying LRCs in growing feather follicles. One-month-old chickens were given drinking water containing 5-bromo-deoxyuridine (BrdU) for 1 week, which labelled all epithelial cells (Fig. 1b, left, and Supplementary Fig. 1c). During the chase period, cells undergoing active proliferation lost their label, but the collar bulge was enriched in label-retaining cells (Fig. 1b, middle, arrow). Some randomly scattered epidermal cells also retained label (arrowhead). By contrast, a 1-h pulse of BrdU labelled transient amplifying cells in the basal layer, excluding the collar bulge region (Fig. 1b, right).

We mapped the multipotency of these cells by measuring how they are incorporated into hosts. The first assay involved quail–chicken grafting¹². We microdissected adult quail follicle epithelium into several fragments (Fig. 1c, left, and Supplementary Fig. 2a) and transplanted them into chicken limb buds at embryonic day 5 (E5).

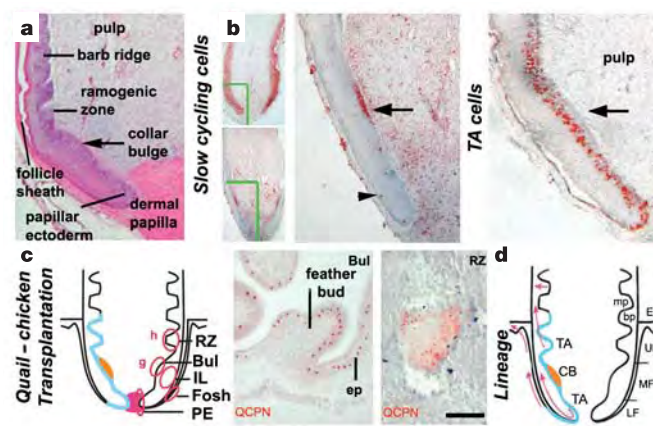


Figure 1 | Identification of feather epithelial stem cells. **a**, Feather follicle. Half a feather follicle is shown with haematoxylin and eosin staining. See Supplementary Fig. 1a, b for a more complete illustration. **b**, Left, BrdU labelling before (top) and 1 week after (bottom) chase in the whole follicle and collar region. Middle, enlargement of boxed area in bottom left image. Enlargement of boxed area in top left image is shown in Supplementary Information. Right, a 1 h pulse of BrdU leaves the collar bulge unlabelled. Arrows indicate LRC; arrowhead indicates scattered LRCs. **c**, Left, regions dissected from adult quail feather follicles for grafting to chicken embryos (see Supplementary Fig. 2 for microdissection and Table 1 for results). Collar bulge (Bul) transplant after 5 d (middle) shows that quail cells (red; stained with QCPN antibody) are incorporated into bud, interbud and barb ridges. Transplants from the ramogenic zone (RZ; right), papilla ectoderm (PE), intermediate layer (IL) and follicle sheath (Fosh) do not show incorporation (Table 1). Dark grains in right image are carbon particle markers. **d**, Summary of epithelial cell migration in regenerating feather follicles. bp, barb plate; CB, collar bulge; LF, lower follicle sheath; MF, middle follicle sheath; mp, marginal plate; TA, transient amplifying cells; UF, upper follicle sheath. Scale bar, 100 μ m.

¹Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [†]Present address: Department of Mechanical Engineering, University of California Los Angeles, Los Angeles, California 90095, USA.

From the explanted bulge, quail keratinocytes were incorporated into the feather bud, interbud epidermis (Fig. 1c, middle), basal layer of barb ridges and barbule plates, among others (Supplementary Fig. 2b). By contrast, transplants from the ramogenic zone (the region where transient amplifying cells start to differentiate and form branches; see Fig. 1a), intermediate layer (suprabasal layers in the collar), or follicle sheaths did not participate in the formation of host skin (Fig. 1c, right), and the papillar ectoderm showed a low incorporation rate (Table 1). We further verified these results by pre-labelling cells with BrdU before grafting. LRCs migrated out and participated in skin morphogenesis, whereas transient amplifying cells stayed at the site of transplantation (Supplementary Fig. 2c). The second assay used adult follicles as hosts. To avoid immunological rejection, we used DiI to label microdissected follicular fragments, which were autologously transplanted to a different follicle. After 5 d, the bulge-derived cells were incorporated into different structures in the feather filaments, whereas fragments from the ramogenic zone or feather sheath were not (Supplementary Fig. 2d).

We then traced the lineage and fate of feather precursor cells in the collar region with BrdU pulse-chase and DiI labelling (Supplementary Fig. 3). In both, we observed that transient amplifying cells in the lower feather follicle generate cells that became feather filaments, feather sheaths, follicle sheaths and interfollicular epidermis (Fig. 1d). This is consistent with the notion that the collar region contains the source of stem and transient amplifying cells, favouring the previous view that new cells are generated and then specified⁴.

Feather follicles go through cycles of tissue remodelling to moult and to regenerate^{6,13}. We next considered what happens to the stem cells during this process. In the resting phase, the calamus (the feather shaft beneath the vane) keratinizes and the collar is reduced in size. Towards the base, the calamus tapers and eventually the wedge-shaped terminus is dislodged from the rest (Fig. 2a–c). Owing to shrinkage of the collar, the bulge morphology is lost, but LRCs remain and shift towards the lower collar, eventually coming in direct contact with the dermal papilla (Fig. 2a). In the moulting phase, LRCs become nested by the dermal papilla (Fig. 2b and Supplementary Fig. 4), taking the position of the papillar ectoderm in the growth phase follicles (yet the papillar ectoderm in the growth phase is not populated with LRCs). Dislodging of the old feather is followed by regeneration of a plump collar and initiation of new feather growth. The bulge reforms and gradually ascends back to the higher position. LRCs are diffuse in the beginning (Fig. 2c) and then become restricted, reassuming the typical collar bulge configuration in growth phase (see Fig. 1b, middle). Proliferating-cell nuclear antigen staining was used to locate transient amplifying cells in these remodelling follicles. Not much proliferation was found in the moulting phase follicles (Fig. 2, right panels). In the initiation phase, cells, excluding those in the LRC region, started to proliferate.

Table 1 | Multipotency of follicular components in the quail-chicken graft assay

	Transplantation donor	LRC	% incorporation	n
Growth phase	Intermediate layer	–	0	10
	Ramogenic zone	–	0	10
	Collar bulge	+	57	14
	Papillar ectoderm	±	12.5	8
	Follicle sheath	–	0	4
Moulting phase	Calamogenic zone	–	0	4
	Remaining collar (including PE)	+	75	8
	Follicle sheath	–	0	4

Different feather follicular components from growth and moulting phase adult quail feather follicles were dissected (Fig. 1c) and grafted to chicken embryonic limb buds. n indicates number of experiments. Quail feather fragment incorporation into chicken skin is shown as a percentage. The presence of a LRC population is based on data from Figs 1 and 2. In one out of eight experiments, papillar ectoderm (PE) from growth phase follicles shows incorporation.

We then considered how cells in the moulting phase behave in the graft assay. From the resting follicle, we can only isolate the calamus-forming zone, the remaining collar (containing the papillar ectoderm region next to the dermal papilla, populated by LRCs at this stage) and the follicle sheath. Transplantation from quails to chicken hosts showed that cells from the calamogenic zone and the follicle sheath were not incorporated into the host. However, the remaining collar epithelium showed a high efficiency of incorporation into developing skin (Table 1 and Supplementary Fig. 2e). Transplantation of DiI-labelled fragments into adult follicles gave similar results (data not shown). Cells in the follicle sheath labelled with DiI showed no movement in the resting phase follicle and a little bit of movement in the moulting phase follicle, but resumed an upward movement when follicles entered the initial growth phase (Supplementary Fig. 3e).

Because there are different forms of feathers, we considered how LRCs are configured in the growth phase of downy (radially symmetric) and flight (bilaterally symmetric) feather follicles (Fig. 3a). Marginal plates stained for Sonic hedgehog (Shh) were used to help to delineate the orientation of barb ridges in opened adult feather follicle preparations^{14,15} (Fig. 3b). Through helical barb ridge organization, barbs reach the rachis (the feather shaft in the vane) with an angle θ (angle of helical growth)¹⁶. This angle is big in flight feathers, but nearly 0° in downy feathers (Fig. 3b). When feather follicles at the growth phase were sectioned along the anterior–posterior plane, we found that LRC-positive bulges were at the same level along the anterior–posterior axis in downy feathers, but formed a slant (lower on the rachis side) in flight feathers (Fig. 3c). This implies that the LRC ring is tilted towards the rachis side in these feathers. A three-dimensional reconstruction of serial sections¹⁷ indeed shows that LRCs are configured as a horizontally placed ring in adult downy feathers, but a tilted ring in flight feather follicles (Fig. 3d, e).

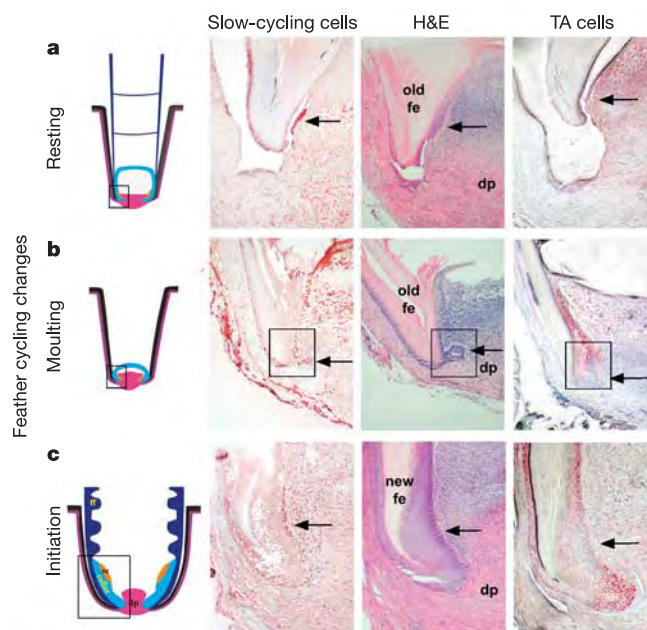


Figure 2 | Stem cells during moulting. LRC and transient amplifying cells were identified in resting (a), moulting (b) and initiation phase (c) follicles. Specimens (right panels) represent boxed regions in the schematic drawings (left panel). Slow-cycling cells (arrow), haematoxylin and eosin stained (H&E) cells, and transient amplifying (TA) cells are shown. Boxed regions from moulting phase follicles are enlarged in Supplementary Fig. 4. Note the changing positions of the LRC niche to accommodate the remodelling of feather follicles. Because stem cells in the feather cycle faster than hairs, LRCs in different stages have to be labelled at different times; therefore, these LRCs may represent the original stem cell population or cells that move into the niche under homeostasis and show stem cell activities. dp, dermal papilla; fe, feather.

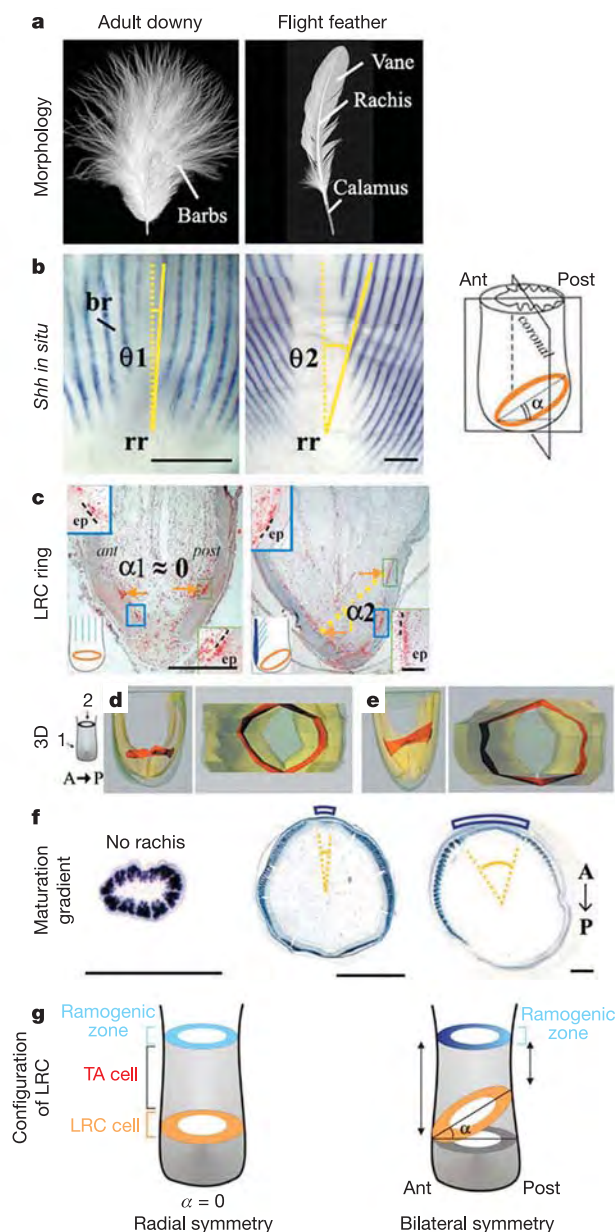


Figure 3 | Configurations of LRC cells in different types of feather. **a**, Gross morphology. **b**, Whole-mount *in situ* hybridization for *Shh* stains marginal plate (regions between barb ridges) to reveal barb ridges (br) and their angle θ with the rachidial ridge (rr). **c**, LRC cells (orange) form a horizontal ring (downy feathers) or an angled ring with a tilt angle α (flight feathers). Lower left insets show drawings for orientation. Also shown are higher magnification views of LRCs in the epithelium (ep, green box) and mesenchyme (blue box). Broken line delineates the basement membrane. Light blue indicates barbs; dark blue indicates rachis. **d**, **e**, Three-dimensional reconstruction of serial sections showing that the LRC ring (red) is horizontal in radially symmetric downy feathers (**d**) but tilted in bilaterally symmetric flight feathers (**e**). The images show side (left) and top (right) views. **f**, Cross-sections at the level of the ramogenic zone. *In situ* hybridization of feather keratin A shows homogenous staining in natal (left) and adult downy (middle) feathers, but an anterior–posterior (A–P) maturation gradient in flight feathers (right). The blue arc and yellow lines delineate the rachis. **g**, Hypothetical model. In bilaterally symmetric flight feathers, the LRC ring forms a tilting angle α . Asymmetry in the LRC region may break the radial symmetry at the ramogenic zone, leading to the formation of a maturation gradient and bilateral symmetry. Scale bars: 100 μ m (**b**); 0.5 mm (**c**, **f**).

The LRC tilting angle α approaches 0° in downy feathers, but increases in flight feathers, reaching a maximum angle of about 45° . We analysed cellular and molecular asymmetries at the ramogenic plane. In embryonic downy feathers, barb ridges form around the same time¹⁸. In flight feathers, new barb ridges are generated sequentially from the barb generative zone in the posterior feather follicle¹⁹. With helical organization towards the anterior follicle, the barb ridges slant to create the rachis¹⁶. This anterior–posterior maturation gradient is demonstrated by the expression of feather keratin A (Fig. 3f). Previous work has shown that bone morphogenetic protein (BMP) and *Shh* are involved in the formation of the barbs and rachis^{13,15}. We now can investigate further whether the tilting LRCs are related to the BMP–*Shh* pathway, the maturation gradient and positioning of the rachis.

Our studies show that although feather and hair follicles arose separately around 155–225 million years ago, they both evolved to have LRCs, transient amplifying cells and differentiating keratinocyte populations. Both evolved a follicular structure with localized growth zones in the proximal end. In both, the dermal papilla, sitting in the follicle base, is essential for the cyclic regeneration, and both share morphogenesis signalling pathways. There are also differences. Feather LRCs are located in the proximal feather follicle and bulge into the mesenchymal pulp. The hair follicle bulge is contiguous with the outer layer of the follicle—the outer root sheath¹⁰. The different topology implies that feathers and hairs use different strategies to preserve their stem cells during moulting. As a consequence, cell migration patterns are also different. In hair follicles, a downward flux of bulge to outer root sheath to hair matrix¹ is followed by the upward flux to form hair filaments²⁰. In feathers, two upstream migrations are observed: collar to ramogenic zone to feather filament, and collar to follicle sheath to interfollicular epithelium. Another difference is that feather LRCs are not as quiescent as hair follicle bulge cells, and the label was diluted after 1 month. This may be due to the faster growth rate of feathers. In other organs, the duration of long-term retention has been shown to depend on different local proliferation dynamics⁹. However, the comparison of hairs and feathers renders support to the concept that LRCs and transient amplifying cells are in transient homeostasis and can populate different niches (such as the bulge or papillar ectoderm) as regulated by the state of the microenvironment through different phases of their growth cycles^{21–23}, and that the bulge stem cell model¹ is of paramount importance to the maintenance and regeneration of episodically renewing organs.

Another feature of the feather follicles is that the topology allows unusual evolutionary possibilities. Early feathers were radially symmetric, lacked a rachis and were composed of equivalent barb branches, similar to the downy feathers found today. A rachis emerged later, making feathers bilaterally symmetric, and set up the basis for further evolution towards flight^{24–26}. It is compelling to speculate that the tilting LRCs may help to break the radial symmetry in the ramogenic plane, setting up the anterior–posterior maturation gradient and hence the bilateral symmetry (Fig. 3g). However, this hypothesis remains to be tested. So, what makes the LRC ring tilt? Swapping dermal papilla between downy and flight feathers showed that both feather symmetry and LRC configuration are determined by the dermal papilla (Z.Y., T.-X.J., R.B.W. and C.-M.C., unpublished data). This implies that feather symmetry is not irreversibly molecularly encoded in specific feather follicles. Rather, feather stem cells are true stem cells that can be moulded into different feather forms, depending on the niches created by the dermal papilla. Thus, here we have answered several decade-old questions about feather stem cells, but have exposed new areas to be explored concerning how different complex forms can be built from epidermal stem cells from one cycle to the next.

METHODS

BrdU labelling. For pulse labelling, chickens were injected intraperitoneally with

BrdU (Sigma) at 50 mg per kg (body weight). Feathers were collected 1 h later. For label-retaining studies of growth phase follicles, 1-month-old chickens were fed BrdU in their drinking water (1 mg ml^{-1}) for 1 week, and 'chased' (left to metabolize the BrdU in their system) for 1 week, 2 weeks, and so on. Samples were fixed in 4% paraformaldehyde, mounted in paraffin, sectioned at $8 \mu\text{m}$, stained with antibody against BrdU (Chemicon) and counterstained with haematoxylin and eosin.

In chickens, neighbouring flight feathers on the wing moult in sequential order. The order helps us to predict the feather cycle stage of a given feather and the timing was verified by vascularity in the calamus region. In our colony, flight feathers moult in chicken aged about 2 months. At 3–4 weeks before the expected events, we fed chickens BrdU for 1 week and chased for 2 weeks. The birds were then killed so that flight feathers at different cycling stages coexisted from the proximal to the distal wing.

DiI labelling. DiI (1,1'-diiododecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate; Molecule Probes) was prepared as described¹¹. We anaesthetized 3-month-old chickens with 10 mg per kg (body weight) of ketamine and xylocaine (2:1). DiI was injected through the follicle wall. Follicles were collected at desired times and processed.

Immunostaining and *in situ* hybridization. Immunostaining and *in situ* hybridizations were processed as described²⁷ with an automated Discovery system (Ventana Medical Systems). RNA probes for feather keratin A (nucleotides 571–1099; X17511) have been described²⁸.

Multipotentiality grafting assay. For quail–chicken grafting, different parts of quail feather follicles were dissected (Fig. 1c and Supplementary Fig. 2a) and transplanted to E5 chicken embryo limb buds. At various stages of development, embryos were collected, fixed and sectioned. Samples were stained for quail-cell-specific antigen (QCPN antibody, Hybridoma Bank).

For adult follicle transplantation, similar dissections were done, labelled with DiI and then transplanted back to another feather follicle in the same chicken. Follicles were collected after 4–5 d.

Three-dimensional reconstruction. We used ten sections ($42 \mu\text{m}$ between each) for reconstruction²⁹. Sections were digitized using IGL Trace software (<http://synapses.bu.edu/tools/trace/trace.htm>). These images were aligned and rendered as a three-dimensional view of LRC using Rhinoceros NURBS modelling for Windows (Robert McNeel and Associates).

Received 15 December 2004; accepted 12 September 2005.

- Morris, R. J. *et al.* Capturing and profiling adult hair follicle stem cells. *Nature Biotechnol.* **22**, 411–417 (2004).
- Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K. & Barrandon, Y. Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* **104**, 233–245 (2001).
- Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
- Lillie, F. R. & Wang, H. Physiology of development of the feather. V. Experimental morphogenesis. *Physiol. Zool.* **14**, 103–133 (1941).
- Lillie, F. R. & Wang, H. Physiology of development of the feather. VII. An experimental study of induction. *Physiol. Zool.* **17**, 1–32 (1944).
- Lucas, A. M. & Stettenheim, P. R. (eds) *Avian Anatomy: Integument. Agriculture Handbook 362: Agricultural Research Services* (US Dept. Agriculture, Washington DC, 1972).
- Wang, H. The morphogenetic function of the epidermal and dermal components of the papilla in feather regeneration. *Physiol. Zool.* **16**, 325–350 (1943).
- Cohen, J. & Espinasse, P. G. On the normal and abnormal development of the feather. *J. Embryol. Exp. Morphol.* **9**, 223–251 (1961).
- Potten, C. S., Owen, G. & Booth, D. Intestinal stem cells protect their genome by selective segregation of template DNA strands. *J. Cell Sci.* **115**, 2381–2388 (2002).
- Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61**, 1329–1337 (1990).
- Bronner-Fraser, M. Manipulations of neural crest cells or their migratory pathways. *Methods Cell Biol.* **51**, 61–79 (1996).
- Dupin, E., Ziller, C. & Le Douarin, N. M. The avian embryo as a model in developmental studies: chimeras and *in vitro* clonal analysis. *Curr. Top. Dev. Biol.* **36**, 1–35 (1998).
- Yu, M. *et al.* Developmental biology of feather follicles. *Int. J. Dev. Biol.* **48**, 181–191 (2004).
- Ting-Berrett, S. A. & Chuong, C. M. Sonic hedgehog in feather morphogenesis: Induction of mesenchymal condensation and association with cell death. *Dev. Dyn.* **207**, 157–170 (1996).
- Harris, M. P., Fallon, J. F. & Prum, R. O. Shh–Bmp2 signalling module and the evolutionary origin and diversification of feathers. *J. Exp. Zool.* **294**, 160–176 (2002).
- Prum, R. O. & Williamson, S. Theory of the growth and evolution of feather shape. *J. Exp. Zool.* **291**, 30–57 (2001).
- Fiala, J. C. & Harris, K. M. PC-based alignment and reconstruction of serial sections. *Microsc. Anal.* **52**, 5–7 (2002).
- Chuong, C. M. & Edelman, G. M. Expression of cell-adhesion molecules in embryonic induction. I. Morphogenesis of nestling feathers. *J. Cell Biol.* **101**, 1009–1026 (1985).
- Chuong, C. M. & Edelman, G. M. Expression of cell-adhesion molecules in embryonic induction. II. Morphogenesis of adult feathers. *J. Cell Biol.* **101**, 1027–1043 (1985).
- Kopan, R. *et al.* Genetic mosaic analysis indicates that the bulb region of coat hair follicles contains a resident population of several active multipotent epithelial lineage progenitors. *Dev. Biol.* **242**, 44–57 (2002).
- Watt, F. M. & Hogan, B. L. Out of Eden: stem cells and their niches. *Science* **287**, 1427–1430 (2000).
- Fuchs, E., Tumbar, T. & Guasch, G. Socializing with the neighbors: stem cells and their niche. *Cell* **116**, 769–778 (2004).
- Koster, M. I., Huntzinger, K. A. & Roop, D. R. Epidermal differentiation: transgenic/knockout mouse models reveal genes involved in stem cell fate decisions and commitment to differentiation. *J. Invest. Dermatol. Symp. Proc.* **7**, 41–45 (2002).
- Prum, R. O. Development and evolutionary origin of feathers. *J. Exp. Zool.* **285**, 291–306 (1999).
- Yu, M., Wu, P., Widelitz, R. B. & Chuong, C. M. The morphogenesis of feathers. *Nature* **420**, 308–312 (2002).
- Chuong, C.-M. *et al.* Adaptation to the sky: defining the feather with integument fossils from Mesozoic China and experimental evidence from molecular laboratories. *J. Exp. Zool.* **298**, 42–56 (2003).
- Jiang, T.-X., Jung, H. S., Widelitz, R. B. & Chuong, C.-M. Self organization of periodic patterns by dissociated feather mesenchymal cells and the regulation of size, number and spacing of primordia. *Development* **126**, 4997–5009 (1999).
- Chodankar, R. *et al.* Shift of localized growth zones contributes to skin appendage morphogenesis: role of the Wnt/ β -catenin pathway. *J. Invest. Dermatol.* **120**, 20–26 (2003).
- Wu, P., Jiang, T.-X., Suksaweang, S., Widelitz, R. B. & Chuong, C. M. Molecular shaping of the beak. *Science* **305**, 1465–1466 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Cotsarelis and R. Prum for helpful input; and all members of the Chuong laboratory for discussion. This work is supported by grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (to C.M.C.) and from the National Cancer Institute (to R.B.W.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.M.C. (chuong@pathfinder.usc.edu).

LETTERS

The zebrafish dorsal axis is apparent at the four-cell stage

Aniket V. Gore^{1,2}, Shingo Maegawa³, Albert Cheong¹, Patrick C. Gilligan¹, Eric S. Weinberg³
& Karuna Sampath^{1,2}

A central question in the development of multicellular organisms pertains to the timing and mechanisms of specification of the embryonic axes. In many organisms, specification of the dorsoventral axis requires signalling by proteins of the Transforming growth factor- β and Wnt families^{1–3}. Here we show that maternal transcripts of the zebrafish Nodal-related morphogen, Squint (Sqt), can localize to two blastomeres at the four-cell stage and predict the dorsal axis. Removal of cells containing *sqt* transcripts from four-to-eight-cell embryos or injection of antisense morpholino oligonucleotides targeting *sqt* into oocytes can cause a loss of dorsal structures. Localization of *sqt* transcripts is independent of maternal Wnt pathway function and requires a highly conserved sequence in the 3' untranslated region. Thus, the dorsoventral axis is apparent by early cleavage stages and may require the maternally encoded morphogen Sqt and its associated factors. Because the 3' untranslated region of the human *nodal* gene can also localize exogenous sequences to dorsal cells, this mechanism may be evolutionarily conserved.

Localized maternal factors specify the embryonic axes in several organisms⁴. In many vertebrates, specification of the dorsoventral (DV) axis is thought to require maternally encoded signals of the Transforming growth factor (TGF)- β and Wnt pathways^{5,6}. In *Xenopus*, on the entry of sperm, cortical rotation results in the transport of dorsal determinants by means of microtubules⁷. In fish, removal of the vegetal yolk cell and disruption of microtubules by cold shock or by drugs disrupts the movement of dorsal determinants from the yolk to the blastoderm^{8–10}. However, the nature of these dorsal determinants and the mechanism of movement are poorly understood.

In zebrafish, transcripts of the diffusible morphogen Sqt are expressed in oocytes and embryos^{11–14}. Although *sqt* RNA is present uniformly through oogenesis, on egg activation and fertilization it rapidly localizes to the blastoderm by a microtubule-dependent process¹⁴. Subsequently, during early gastrula stages, before the formation of the zebrafish organizer, the embryonic shield, *sqt* RNA is expressed in the dorsal side of the embryo^{11–13}.

We find, by *in situ* hybridization, that *sqt* RNA is localized asymmetrically in four-cell and eight-cell embryos (66%, $n = 72$, and 67%, $n = 52$, respectively; Fig. 1a–c, e–g, and Supplementary Table 1), and is detectable at least until the 32- and 64-cell stages (Fig. 1d, h). Real-time reverse-transcriptase-mediated polymerase chain reaction (RT–PCR) analysis (Supplementary Fig. S1) shows that *sqt* RNA levels are constant through the cleavage and blastula stages. To examine the dynamics of *sqt* RNA localization in living embryos, we injected synthetic Alexa-488-labelled *sqt* RNA into one-cell embryos and observed the movement of the fluorescent RNA by time-lapse videomicroscopy. Injected fluorescent *sqt* RNA is detected in the cells expressing endogenous *sqt* RNA (72%, $n = 19$; Fig. 1m–o),

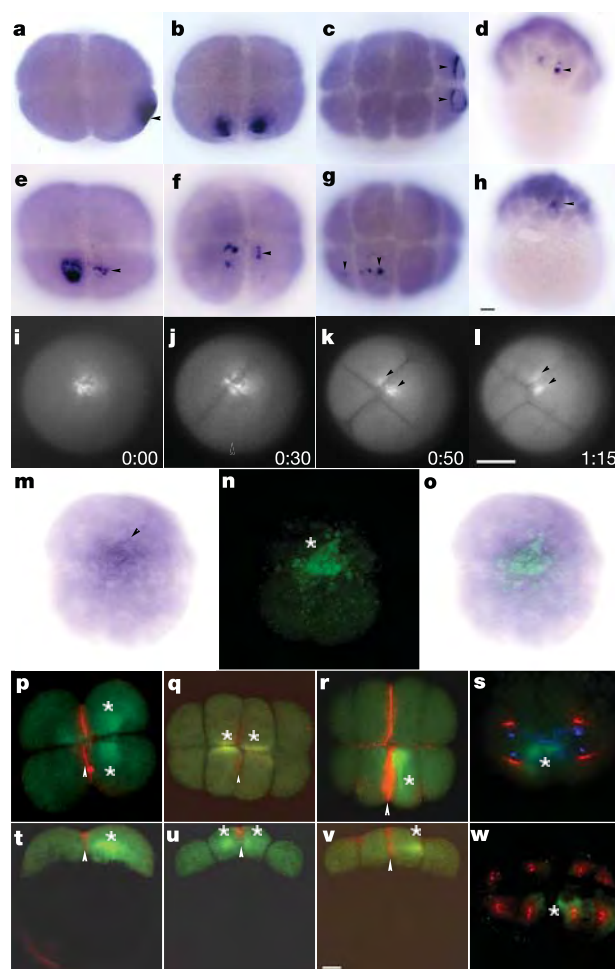


Figure 1 | Sqt RNA localizes in early embryos. **a–h**, *In situ* hybridization to detect endogenous *sqt* RNA (arrowheads) in embryos at the 4-cell (**a**, **b**, **e**, **f**), 8-cell (**c**, **g**), 32-cell (**d**) and 64-cell (**h**) stages. **i–l**, Stills from a movie of an embryo injected with Alexa-488-labelled *sqt* RNA (arrowheads) at the one-cell (**i**), two-cell (**j**), four-cell (**k**) and eight-cell (**l**) stages. Times are indicated as h:min. **m–o**, Four-cell embryo showing localization of endogenous *sqt* RNA (arrowhead, **m**), injected fluorescent *sqt* RNA (asterisk, **n**), and their overlap (**o**). **p–r**, First cleavage furrow labelled with FM4-64 (red; arrowhead) and Alexa-488-labelled *sqt* RNA (green; asterisks), in four-cell (**p**, **t**) and eight-cell (**q**, **r**, **u**, **v**) embryos. **s**, **w**, Immunodetection of γ -tubulin (red), DNA (blue) Alexa-488-labelled *sqt* RNA (green; asterisks) in four-cell (**s**) and eight-cell (**w**) embryos. **a–c**, **e–g**, **i–l**, **m–s**, **w**, animal pole views; **d**, **h**, **t–v**, lateral views. Scale bar, 100 μ m in **h**, **l** and **v**, for **a–h**, **i–l** and **p–w**, respectively.

¹Vertebrate Development Group, Temasek Life Sciences Laboratory, 1 Research Link, 117604 Singapore. ²Department of Biological Sciences, National University of Singapore, 117543 Singapore. ³Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

and is asymmetrically localized in two cells of four-cell embryos ($n = 20$; Fig. 1i–l, and Supplementary movies 1 and 2). In contrast, control RNAs encoding β -galactosidase or green fluorescent protein (GFP; $n = 23$; Supplementary movie 3), or the Nodal co-receptor *Oep*⁶ (data not shown), are distributed uniformly in all blastomeres. Furthermore, injected fluorescent *sqt* RNA fails to localize to the blastoderm in embryos treated with nocodazole, in a similar manner to endogenous *sqt* RNA¹⁴ ($n = 17$; Supplementary movie 4). We therefore conclude that injected fluorescent *sqt* RNA is localized at the four-cell stage in a similar manner to endogenous *sqt* RNA and requires microtubules for localization.

In *Xenopus*, the first cleavage plane typically demarcates the DV axis¹⁵. Because *sqt* RNA can localize asymmetrically in four-cell embryos, we tested whether the localization correlates with the initial cleavage planes by labelling the first cleavage furrow with the fluorescent marker FM4-64 (ref. 16). Embryos injected with fluorescent *sqt* RNA show injected *sqt* RNA in cells on one side ($n = 31$; Fig. 1p, r, t, v) or on both sides of the cleavage furrow ($n = 20$ embryos; Fig. 1q, u). Similar results were obtained when the second and third cleavage planes were labelled ($n = 21$, data not shown). This is supported by expression of endogenous *sqt* RNA at the eight-cell stage (Fig. 1c, g). Localization of *sqt* RNA is therefore random with respect to the early cleavage planes.

In the mollusc *Ilyanassa obsoleta*, RNA encoding the secreted morphogen Decapentaplegic (Dpp) is localized in four macromere cells at the eight-cell stage¹⁷. Furthermore, *Iodpp* RNA localizes to the pericentriolar matrix in interphase centrosomes and is partitioned to one daughter cell during cell division¹⁷. We examined whether *sqt* RNA localizes to the centrosomes in four-cell and eight-cell embryos. Localized *sqt* RNA does not overlap with γ -tubulin-stained centrosomes in 90% of embryos examined ($n = 30$; Fig. 1s, w). We detected *sqt* RNA overlapping with one of the centrosomes in only three embryos (not shown). We conclude that *sqt* RNA does not localize to the centrosomes.

Localization of several transcripts in *Drosophila* and *Xenopus* embryos and in mammalian neurons requires elements in the 5' and 3' untranslated regions (UTRs)⁴. To identify the element required for localization of *sqt* RNA, we generated deletions in its UTRs. Deletion of the *sqt* 3' UTR ($n = 8$; Fig. 2a–c and Supplementary movie 5) or fusion of *sqt* coding sequences with the 3' UTR of the β -globin gene results in uniform distribution of injected fluorescent RNA ($n = 17$; Fig. 2d–f, and Supplementary movie 6). In contrast, injected *lacZ* RNA fused with the *sqt* 3' UTR is localized ($n = 21$; Fig. 2g, h, and Supplementary movies 7 and 8). Furthermore, β -Gal protein co-localizes with the early markers of the dorsal yolk syncytial layer and dorsal blastomeres in zebrafish^{18,19}, *dharmabozozok* (67%,

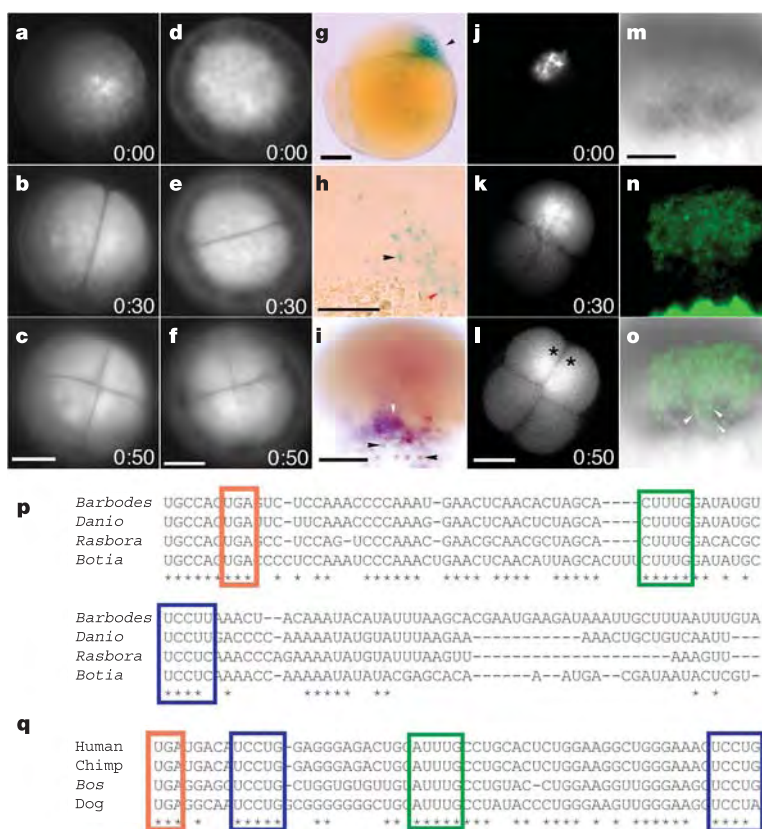


Figure 2 | The *sqt* 3' UTR is necessary and sufficient for dorsal localization. **a–c**, Stills from a movie of an embryo injected with *sqt* RNA with deletions in the 3' UTR. Times are indicated as h:min. **d–f**, Stills from a movie of *sqt* RNA with the 3' UTR of β -globin gene. **g–i**, β -Galactosidase (**g**, **h**) staining in the blastoderm (black arrowhead in **g**, **h**) and yolk syncytial nuclei (red arrowhead in **h**) of 1,000-cell-stage embryos injected with *lacZ:sqt* 3' UTR RNA, and co-localization with *dharmabozozok* expression (white arrowhead in **i**; black arrowheads indicate nuclear localization of β -galactosidase). **j–l**, Stills from a movie of embryo injected with *lacZ:human nodal* 3' UTR RNA (asterisks in **l**) at the one-cell (**j**), two-cell (**k**) and four-cell (**l**) stages. **m–o**, *In situ* hybridization to detect *dharmaboz* (**m**), nuclear β -Gal (**n**) and their co-localization (arrowheads, **o**), in embryos injected with *lacZ:human nodal* 3' UTR RNA. **p**, Alignment of the first 80

nucleotides of the *sqt* 3' UTR from *Barbodes gonionotus* (DQ080242, nucleotides 338–449), *Danio rerio* (NM_130966, nucleotides 1219–1317), *Rasbora heteromorpha* (DQ080243, nucleotides 266–355) and *Botia macracantha* (DQ080241, nucleotides 352–459). The stop codon is boxed in red, and two prominent conserved blocks are boxed in green and blue. **q**, Alignment of the first 60 nucleotides of the *nodal* 3' UTR from *Homo sapiens* (NM_018055), *Pan troglodytes* (NW_113666.1, nucleotides 33040–33099), *Bos taurus* (NW_930639.1, nucleotides 21489–21547) and *Canis familiaris* (NW_876311.1, nucleotides 11084132–11084192). Note two motifs, (C/A)UUUG (boxed in green) and UCCU(G/A) (boxed in blue), shared between the cyprinid and mammalian alignments. **a–f**, **j–l**, Animal pole views; **g–i**, **m–o**, lateral views. Scale bar, 100 μ m in **c**, **f**, **g–i** and **l** for **a–l**, and 25 μ m in **o** for **m–o**.

$n = 52$; Fig. 2i) and nuclear β -catenin ($n = 12$, data not shown). The *sqt* 3' UTR is therefore necessary and sufficient for localization of *sqt* RNA to the future dorsal blastomeres.

To define the localization element further, we performed systematic deletion analysis in the *sqt* 3' UTR. We observed that the first 50 nucleotides of the *sqt* 3' UTR are sufficient to localize *sqt* RNA by the four-cell stage ($n = 16$; Supplementary movies 9 and 10). On amplification of the 3' UTR of *sqt* genes from three cyprinid fish,

Rasbora heteromorpha, *Barbodes gonionotus* and *Botia macracantha*, we found several conserved sequences in the alignment, including one that contains a highly conserved CUUUG motif (green box in Fig. 2p). Part of the 3' UTR of the *nodal* genes is also conserved between human, chimpanzee, cow and dog (<http://genome.ucsc.edu>) (Fig. 2q). Within the mammalian alignment there are conserved blocks with significant similarity to the motifs identified in the cyprinid alignment. To test whether the mammalian element is able to confer dorsal

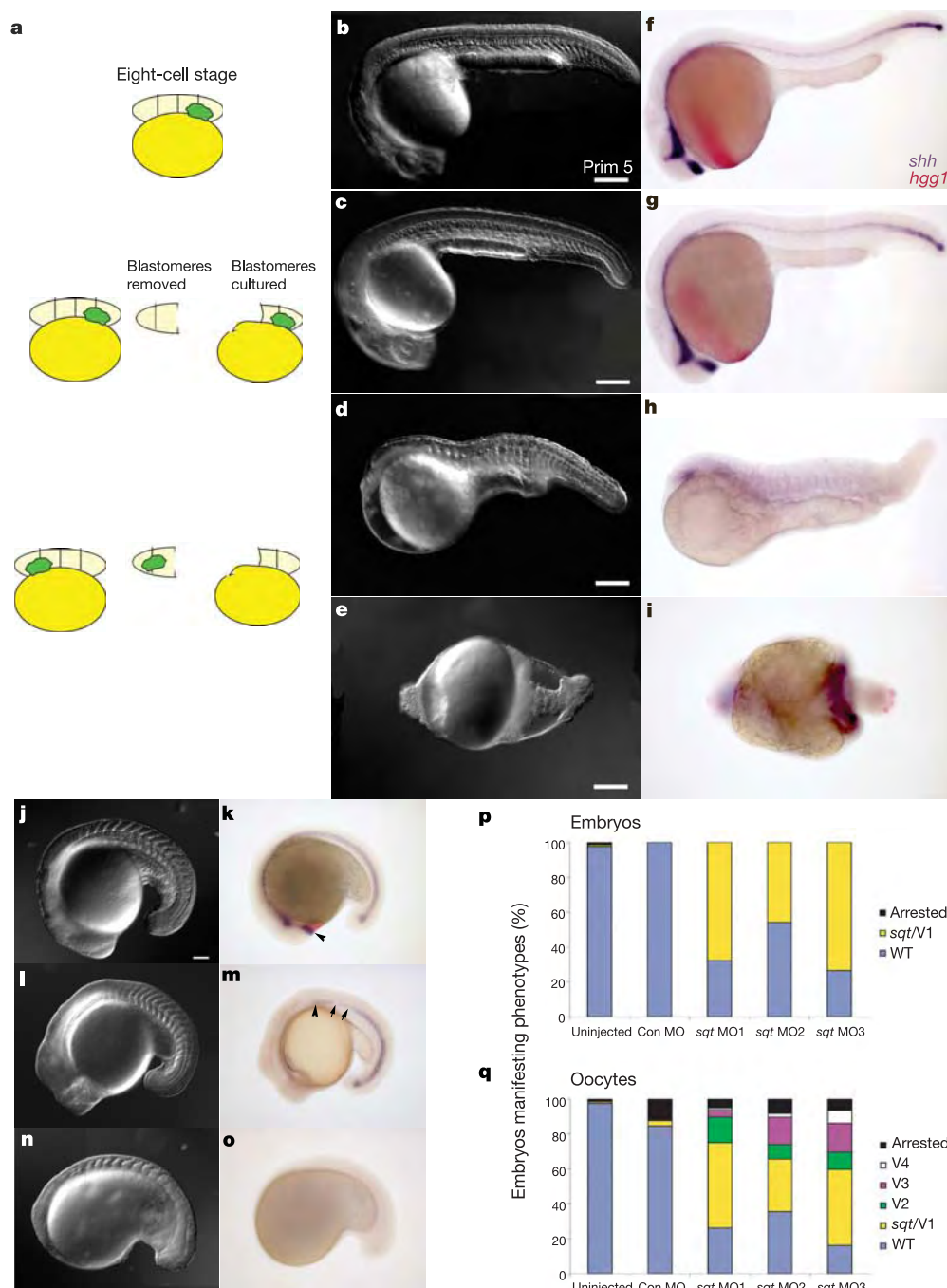


Figure 3 | Removal of cells with *sqt* RNA from embryos or injection of *sqt* morpholinos into oocytes leads to loss of dorsal structures. **a**, Schematic representation of eight-cell embryos to show removal of blastomeres with Alexa-488-labelled *lacZ:sqt* 3' UTR RNA (green, *sqt*⁺) or control removal of cells without *sqt* (*sqt*⁻). **b–e**, Live embryos at 24 h showing wild-type (WT) phenotypes in unmanipulated control (**b**), *sqt*⁻ cell removed (**c**), or *sqt*⁺ removed (**d**, **e**) embryos. **f–i**, Expression of *shh* (purple) and *hgg* (red) in unmanipulated (**f**), control *sqt*⁻ removed (**g**) or *sqt*⁺ removed (**h**, **i**) embryos. Phenotypes range from mild ventralization, V2 (not shown), and

loss of anterior structures, V3 (**d**, **h**), to complete ventralization, V4 (**e**, **i**), on *sqt*⁺ cell removal. Injection of morpholinos (MOs) against *sqt* into oocytes (**n**, **o**), also results in loss of dorsal and anterior structures, in comparison with oocytes injected with control *sqt* mismatch morpholinos (**j**, **k**, **q**) or fertilized embryos injected with *sqt* morpholinos (**l**, **m**, **p**). **j**, **l**, **n**, Live embryos at 18 h; **k**, **m**, **o**, expression of *shh* and *hgg* in the same embryos. Arrowheads in **k** and **m** mark the anterior limit of *shh* expression, and arrows in **m** indicate gaps in *shh* expression. Lateral views with anterior to the left. Scale bars, 100 μ m (**b**, **j**) and 50 μ m (**c–e**).

localization, we injected *lacZ*:human *nodal* 3' UTR fusion RNA into zebrafish embryos at the one-cell stage. Fluorescent *lacZ*:*hsnodal* 3' UTR RNA is localized in two cells of four-cell embryos, in a similar manner to *sqt* RNA ($n = 11$; Fig. 2j–l, and Supplementary movies 11 and 12). Furthermore, β -Gal protein is expressed in cells expressing *dharma/boz* (72%, $n = 18$; Fig. 2m–o) and nuclear β -catenin (67%, $n = 28$; Supplementary Fig. S2), indicating that the human *nodal* 3' UTR localizes β -Gal to dorsal cells in zebrafish embryos. This shows that, in a similar manner to *sqt*, the mammalian *nodal* 3' UTR can also localize heterologous transcripts to the future dorsal side of the zebrafish embryo.

Because *sqt* RNA can asymmetrically localize to the prospective dorsal cells by the four-cell to eight-cell stage, we investigated whether the dorsal axis is specified by these stages. We injected fluorescent *lacZ*:*sqt* 3' UTR RNA to mark the *sqt*-containing cells in early embryos and removed these blastomeres (*sqt*⁺) in four-to-eight-cell embryos. As controls, we removed blastomeres without *sqt* (*sqt*[−]) (Fig. 3a). Twelve of 27 (44%) embryos in which *sqt*⁺ cells were removed showed various degrees of ventralization, ranging from cyclopia and breaks in the notochord (not shown) to loss of dorsal and anterior structures and complete ventralization²⁰ (Fig. 3d, e, h, i, and Supplementary Table 2). In contrast, if the *sqt*⁺ cells were retained, 95% of the operated embryos were morphologically normal, although small ($n = 19$; Fig. 3c, g). Because most control embryos are normal (Fig. 3b, c, f, g), and localization of *sqt* RNA is dynamic through early cleavage stages, the *sqt*⁺ cells may be required for dorsal structures. To confirm this, we also analysed *sonic hedgehog* (*shh*) expression in the axial midline, and that of the hatching gland marker gene *hgg*, in the prechordal plate mesoderm of the operated embryos. Whereas all control dissected embryos expressed both genes ($n = 19$; Fig. 3g), 48% of embryos in which *sqt*⁺ cells were removed had severely diminished or no *shh* expression (Fig. 3h, i) and did not express *hgg* ($n = 27$; Fig. 3h, i, and Supplementary Table 2). Dorsal specification is therefore initiated as early as cleavage stages in zebrafish, with the cells containing localized maternal *sqt* RNA required for the formation of dorsal structures.

If maternal *sqt* RNA is required for dorsal specification, injection of morpholinos should result in severely ventralized phenotypes. However, embryos injected with *sqt* morpholinos are similar to zygotic *sqt* mutant embryos²¹ (73%, $n = 174$ for *sqt* MO3; Supplementary Table 3), with anterior cyclopia (Fig. 3l, m, p) and breaks or gaps in the notochord (Fig. 3m, black arrows). Recent studies in *Xenopus* have shown that injection of antisense oligonucleotides targeting maternal transcripts into oocytes results in more severe phenotypes than that seen on injection into fertilized embryos²². To test whether maternal *sqt* is required for dorsal specification, we injected three non-overlapping antisense morpholino oligonucleotides targeting *sqt* into zebrafish oocytes. In all, 77% of fertilized embryos from morpholino-injected oocytes ($n = 272$ for *sqt* MO3) manifest phenotypes, of which 44% are more severe (Fig. 3n, o, q) than that seen on injection of *sqt* morpholinos into fertilized embryos ($n = 174$; Fig. 3l, m, p)²¹ or *sqt* and *MZsqt* mutant embryos^{6,13}. We observed these phenotypes with three non-overlapping morpholinos towards *sqt* but not with control mismatch morpholinos (Fig. 3j, k, q and Supplementary Table 3). These results suggest a role for maternal *sqt* in the specification of dorsal structures.

Maternal Wnt signalling is important for organizer formation and dorsal specification in frogs and fish^{2,7,23}. Embryos from mothers homozygous for mutations in the maternal effect gene *ichabod* (*ich*) show dorsal deficiencies and fail to localize β -catenin in dorsal nuclei²³. This dorsal deficit and lack of dorsal nuclear β -catenin is due to an impairment of maternal expression of a second β -catenin gene, *β -catenin2*, which maps in proximity to the *ich* mutation and is required for establishment of the early dorsal signalling centre (G. Bellipanni, M. Varga, S.M., Y. Imai, C. Kelly, A. Pomrehn, F. Chu, W. S. Talbot and E.S.W., unpublished observations). To determine

whether *sqt* RNA localization is dependent on β -catenin function, we injected fluorescent *sqt* RNA into embryos from homozygous *ich* mutant mothers. The fluorescent *sqt* RNA is localized (Fig. 4a–f) in *ich* mutant embryos. Localization of maternal *sqt* RNA is therefore either upstream of or parallel to the β -catenin pathway. Furthermore, injection of full-length *sqt* RNA with its 3' UTR in *ich* embryos results in rescue of the mutant phenotypes (Fig. 4g, h, l)²³. Whereas more than 60% of uninjected *ich* mutant embryos are completely ventralized, and nearly 40% are severely ventralized ($n = 42$; Fig. 4g, l), fewer than 20% of *sqt*⁺ RNA-injected *ich* embryos are strongly ventralized, and 13% show complete rescue of the mutant phenotypes ($n = 36$; Fig. 4h, l). Consistent with the zygotic function of *sqt* is our observation that injected *sqt* RNA rescues dorsal structures in *ich* mutant embryos but does not rescue nuclear accumulation of β -catenin (Fig. 4i–k). Localization of maternal *sqt* RNA is therefore independent of β -catenin.

Our finding that *sqt* RNA can localize by early cleavage stages in cells fated to become dorsal, and that the removal of which can cause ventralization, indicates that the dorsal axis may be specified very

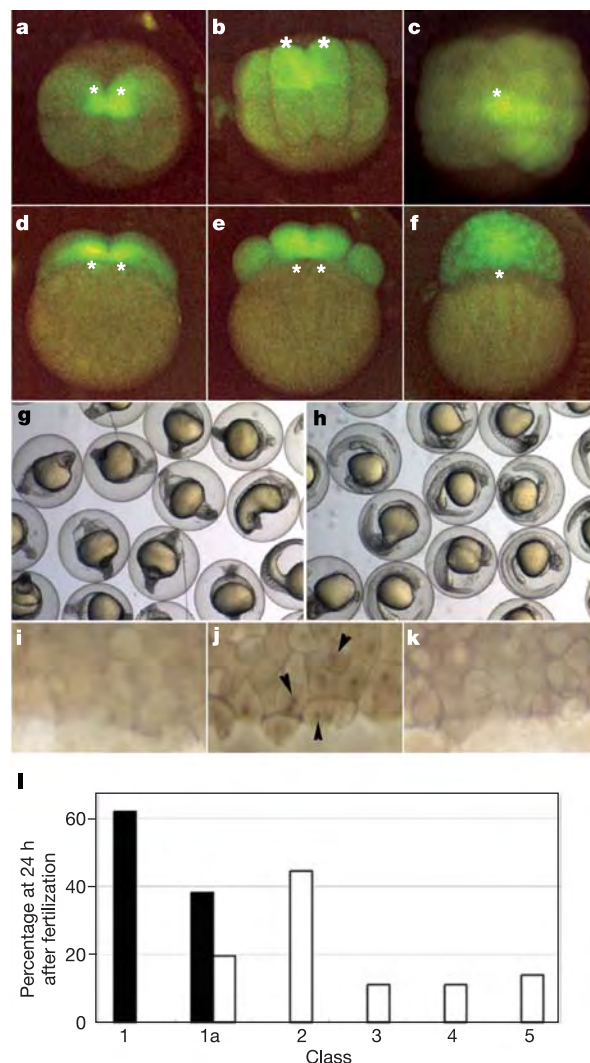


Figure 4 | Localization and function of *sqt* RNA is independent of β -catenin2 function. a–f, Alexa-488-labelled *sqt* RNA (asterisks) is localized in *ich* mutant embryos at the 4-cell (a, d), 8-cell (b, e), 16-cell (c) and 128-cell (f) stages. g, h, l, Rescue of *ich* mutant embryos injected with *sqt* RNA (h; open bars in l; $n = 36$), compared with uninjected *ich* embryos (g; solid bars in l; $n = 42$). i–k, Nuclear β -catenin localization is absent in uninjected *ich* embryos (i), and in embryos injected with *sqt* RNA (k), but present in *ich* mutant embryos injected with β -catenin (black arrowheads in j).

early in zebrafish. This is consistent with ultraviolet irradiation experiments that resulted in embryos with normal epiboly but no dorsal structures⁸. Although *sqt* RNA localizes during cleavage stages in prospective dorsal cells, *sqt* RNA may be localized to any two adjacent cells of four-cell and eight-cell embryos. This confirms previous suggestions that the dorsal axis in zebrafish is independent of the initial cleavage planes^{24,25}.

We also find some embryos with uniform distribution of *sqt* RNA during early cleavage stages, and the proportion of embryos with localized *sqt* RNA increases with development. It is not clear how *sqt* expression is restricted or how dorsal may be specified in these embryos. Our real-time PCR experiments show that levels of *sqt* RNA are constant through early cleavage stages. Although it is possible that the asymmetric distribution of *sqt* RNA is due to active localization, at present we cannot rule out some degradation of unlocalized *sqt* RNA. Removal of the cells with localized *sqt* RNA can cause ventralization, a phenotype more severe than that seen in *sqt* or other nodal signalling mutant embryos⁶. This is probably because the removal of the *sqt*⁺ cells removes maternal Sqt, and possibly other factors, that may be required for dorsal specification. In support of this, injection of antisense morpholinos targeting *sqt* into oocytes can cause severe loss of dorsal structures. Some operated embryos also show anterior–posterior truncations. These phenotypes are not seen in control operated embryos in which *sqt*[−] cells were removed, indicating that *sqt*⁺ cells may be required for many aspects of axis specification. In the goldfish *Carassius auratus*, bisection of two-cell embryos leads to a loss of dorsal and anterior structures in more than 50% of the bisected embryos, whereas only 45% of the bisected pairs develop normally⁹. This indicates that maternal determinants (such as Sqt) may be distributed asymmetrically in this fish as well.

Our finding that the *sqt* dorsal localization element is present in mammalian *nodal* genes and can localize exogenous sequences to dorsal cells in zebrafish raises the possibility that dorsal may be specified or facilitated by a localized *nodal*-containing complex in higher vertebrates as well. This is particularly surprising because mammalian embryos are thought to undergo ‘regulative’ development. It is thought that cell fates are plastic in the early mammalian embryo, with each cell capable of regulating its development to produce a complete organism, and the ability of mouse embryos to recover after the removal of cells from the early embryo supports this notion. In contrast, recent experiments indicate that embryonic polarity may originate very early in mouse development (reviewed in ref. 26). However, the mechanisms are unclear. Examination of early localization of *nodal* in early mammalian embryos could reveal conserved pathways for dorsal specification. Given that the mechanisms that pattern early embryos and numerous other developmental processes are evolutionarily conserved, it is possible that DV axis specification by asymmetric localization of dorsal determinants is conserved in higher vertebrates as well.

METHODS

RNA synthesis and injections. The *squint* complementary DNA was amplified by RT-PCR from whole-ovary RNA, and cloned in the plasmid pCS2 + . Alexa 488 UTP or fluorescein UTP (Molecular Probes) was used to synthesize fluorescent RNA as described²⁷. Fluorescent RNAs typically contained 1 fluorochrome per 250 nucleotides for Alexa-488-labelled transcripts and 1 fluorochrome per 150 nucleotides for fluorescein-labelled transcripts. Aliquots (15 µg) of labelled RNA were injected in 4-nl volumes into wild-type or *ich* embryos²³ at the one-cell stage.

In situ hybridizations. To detect endogenous *sqt* RNA in wild-type embryos, *in situ* hybridization was performed with a full-length *sqt* cDNA antisense probe as described²⁸, and the colour reaction was allowed to proceed overnight. To detect endogenous and injected *sqt* RNA in the same embryos, *in situ* hybridization was performed with a *sqt* coding sequence probe, and injected fluorescent *lacZ:sqt* fusion RNA was detected with anti-fluorescein antibodies.

Furrow labelling. Embryos injected with Alexa-488-labelled *sqt* RNA were dechorionated and incubated in 2 µM FM4-64 (Molecular Probes) in embryo medium as described¹⁶. Embryos were mounted in 2.5% methylcellulose and

optical sections were obtained at 7-µm intervals on a Zeiss Meta 510 Laser scanning microscope.

Immunohistochemistry. For the detection of centrosomes, a rabbit polyclonal antibody generated against γ-tubulin (Sigma) was used. To detect β-catenin, monoclonal antibodies, and for β-galactosidase, rabbit polyclonal antibodies (Chemicon) were used, followed by incubations with anti-mouse Alexa 488 or anti-rabbit Alexa 568 secondary antibodies. β-Galactosidase enzymatic activity was detected as described²⁹. For detection of β-galactosidase protein after *in situ* hybridizations, embryos were incubated with a rabbit polyclonal antibody against β-galactosidase, and horseradish peroxidase staining was carried out with the ABC kit (Pierce). To detect injected fluorescent *sqt* RNA, we used rabbit anti-fluorescein antibodies conjugated with Alexa 488 (Molecular Probes).

Time-lapse video-microscopy. Embryos were injected at the one-cell stage and dechorionated manually. For animal pole views, embryos were mounted on depression slides in 0.75% low-melting-point agarose in embryo medium. A Zeiss Axioplan2 microscope with a Coolsnap HQ camera (Roper Scientific) was used to acquire images. For lateral views, embryos were mounted on cover-slip-bottomed dishes and images were acquired on a Zeiss Axiovert200M inverted microscope with a Coolsnap HQ camera. The Metamorph (Universal Imaging Corporation) and NIH ImageJ software packages were used to acquire and process the images.

Embryo dissections. Wild-type zebrafish embryos injected with Alexa-488-labelled *lacZ:sqt* 3′ UTR RNA were dechorionated and placed in agarose-coated dishes containing 1 × Ringer’s, 1.5% egg albumin and antibiotics⁹. Using an eyelash tool, *sqt*⁺ cells (or *sqt*[−] cells in controls) were removed at the four-cell or eight-cell stage. Operated embryos were left to heal at 28 °C. At the 1,000-cell stage, operated embryos were transferred to 1/3 × Ringer’s containing antibiotics, and incubated at 28 °C overnight. Survival rates ranged from 40% to 75% in eight experiments, with higher survival of embryos dissected at the eight-cell stage.

Morpholino injections. Mature oocytes squeezed from wild-type females were injected with 10-ng aliquots of the following three independent morpholinos towards *squint* (*sqt* MO1, 5′-CAGGAGCCCGCAGGAAAACATGTCA-3′; *sqt* MO2, 5′-ATCTGAGAGATTCTTACCTGCATGT-3′; *sqt* MO3, 5′-ATCAATTG-TACTTACTTTTAGCGAC-3′), or a control *squint* mismatch morpholino (Con MO, 5′-CAGGATCCTGCACGAAAACGTTGTCA-3′; mismatches are underlined), mixed with rhodamine dextran. Oocytes were fertilized *in vitro* with fresh sperm, and fertilized embryos that were positive for rhodamine dextran were sorted and incubated at 28 °C in embryo medium containing antibiotics. For embryo injections, fertilizations were performed before injection.

Cyprinid *sqt* fragments. The tinfoil barb (*Barbodes gonionotus*), harlequin rasbora (*Rasbora heteromorpha*) and clown loach (*Botia macracantha*) were obtained from a local fish shop, and genomic DNA was prepared by standard protocols. Primers (SqtCF1, 5′-AGAAGGTGGATATGTGGGTGGACT-3′; SqtEifCR1, 5′-ATGCCAGTTTGAATGGACATC-3′) were designed with primer design software (<http://scitools.idtdna.com/Primerquest/>) to sequences conserved between the zebrafish and *Fugu* genomes (<http://genome.ucsc.edu/>). PCR was performed and the amplified fragments were gel-purified and sequenced. Sequences were aligned with CLUSTALX³⁰.

Received 4 July; accepted 30 August 2005.

1. Moon, R. T. & Kimelman, D. From cortical rotation to organizer gene expression: toward a molecular explanation of axis specification in *Xenopus*. *BioEssays* **20**, 536–545 (1998).
2. Sokol, S. Y. Wnt signalling and dorso-ventral axis specification in vertebrates. *Curr. Opin. Genet. Dev.* **9**, 405–410 (1999).
3. Holley, S. A. & Ferguson, E. L. Fish are like flies are like frogs: conservation of dorsal–ventral patterning mechanisms. *BioEssays* **19**, 281–284 (1997).
4. Riechmann, V. & Ephrussi, A. Axis formation during *Drosophila* oogenesis. *Curr. Opin. Genet. Dev.* **11**, 374–383 (2001).
5. Heasman, J. Patterning the *Xenopus* blastula. *Development* **124**, 4179–4191 (1997).
6. Schier, A. F. & Talbot, W. S. Molecular genetics of axis formation in zebrafish. *Annu. Rev. Genet.* published online 9 August 2005 (doi:10.1146/annurev.genet.37.110801.143752).
7. Weaver, C. & Kimelman, D. Move it or lose it: axis specification in *Xenopus*. *Development* **131**, 3491–3499 (2004).
8. Jesuthasan, S. & Strahle, U. Dynamic microtubules and specification of the zebrafish embryonic axis. *Curr. Biol.* **7**, 31–42 (1997).
9. Mizuno, T., Yamaha, E. & Yamazaki, F. Localized axis determinant in the early cleavage embryo of goldfish, *Carassius auratus*. *Dev. Growth Differ.* **206**, 389–396 (1997).
10. Oppenheimer, J. The development of isolated blastoderms of *Fundulus heteroclitus*. *J. Exp. Zool.* **72**, 247–269 (1936).
11. Rebagliati, M. R., Toyama, R., Fricke, C., Haffter, P. & Dawid, I. B. Zebrafish *nodal*-related genes are implicated in axial patterning and establishing left–right asymmetry. *Dev. Biol.* **199**, 261–272 (1998).

12. Erter, C. E., Solnica-Krezel, L. & Wright, C. V. Zebrafish *nodal-related 2* encodes an early mesendodermal inducer signalling from the extraembryonic yolk syncytial layer. *Dev. Biol.* **204**, 361–372 (1998).
13. Feldman, B. *et al.* Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* **395**, 181–185 (1998).
14. Gore, A. V. & Sampath, K. Localization of transcripts of the zebrafish morphogen *Squint* is dependent on egg activation and the microtubule cytoskeleton. *Mech. Dev.* **112**, 153–156 (2002).
15. Klein, S. L. The first cleavage furrow demarcates the dorsal–ventral axis in *Xenopus* embryos. *Dev. Biol.* **120**, 299–304 (1987).
16. Feng, B., Schwarz, H. & Jesuthasan, S. Furrow-specific endocytosis during cytokinesis of zebrafish blastomeres. *Exp. Cell Res.* **279**, 14–20 (2002).
17. Lambert, J. D. & Nagy, L. M. Asymmetric inheritance of centrosomally localized mRNAs during embryonic cleavages. *Nature* **420**, 682–686 (2002).
18. Yamanaka, Y. *et al.* A novel homeobox gene, *dharma*, can induce the organizer in a non-cell-autonomous manner. *Genes Dev.* **12**, 2345–2353 (1998).
19. Dougan, S. T., Warga, R. M., Kane, D. A., Schier, A. F. & Talbot, W. S. The role of the zebrafish nodal-related genes *squint* and *cyclops* in patterning of mesendoderm. *Development* **130**, 1837–1851 (2003).
20. Kishimoto, Y., Lee, K. H., Zon, L., Hammerschmidt, M. & Schulte-Merker, S. The molecular nature of zebrafish *swirl*: BMP2 function is essential during early dorsoventral patterning. *Development* **124**, 4457–4466 (1997).
21. Feldman, B. & Stemple, D. L. Morpholino phenocopies of *sqt*, *oep*, and *ntl* mutations. *Genesis* **30**, 175–177 (2001).
22. Tao, Q. *et al.* Maternal *wnt11* activates the canonical *wnt* signalling pathway required for axis formation in *Xenopus* embryos. *Cell* **120**, 857–871 (2005).
23. Kelly, C., Chin, A. J., Leatherman, J. L., Kozlowski, D. J. & Weinberg, E. S. Maternally controlled β -catenin-mediated signalling is required for organizer formation in the zebrafish. *Development* **127**, 3899–3911 (2000).
24. Abdelilah, S., Solnica-Krezel, L., Stainier, D. Y. & Driever, W. Implications for dorsoventral axis determination from the zebrafish mutation *janus*. *Nature* **370**, 468–471 (1994).
25. Helde, K. A., Wilson, E. T., Cretekos, C. J. & Grunwald, D. J. Contribution of early cells to the fate map of the zebrafish gastrula. *Science* **265**, 517–520 (1994).
26. Zernicka-Goetz, M. Patterning of the embryo: the first spatial decisions in the life of a mouse. *Development* **129**, 815–829 (2002).
27. Bullock, S. L. & Ish-Horowicz, D. Conserved signals and machinery for RNA transport in *Drosophila* oogenesis and embryogenesis. *Nature* **414**, 611–616 (2001).
28. Sampath, K. *et al.* Induction of the zebrafish ventral brain and floorplate requires cyclops/nodal signalling. *Nature* **395**, 185–189 (1998).
29. Muller, F., Lele, Z., Varadi, L., Menczel, L. & Orban, L. Efficient transient expression system based on square pulse electroporation and *in vivo* luciferase assay of fertilized fish eggs. *FEBS Lett.* **324**, 27–32 (1993).
30. Higgins, D. G., Bleasby, A. J. & Fuchs, R. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* **8**, 189–191 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Balasubramanian, R. Dosch, R. Dunn, S. Jesuthasan, E. Raz, S. Roy, A. Schier, and members of the Sampath laboratory for discussions and suggestions; S. Koshida, E. Yamaha and H. Takeda for embryo operation techniques; A. Chang and the Qian Hu fish farm for cyprinids; H. Ngoc Bao Quach for technical assistance; M. Hidayat for sequencing; and J. Bradner, C. Hu and C. Goh and the TLL fish facility for fish care. This work was supported by NIH grants (E.W.) and by TLL, Singapore (K.S.).

Author Information The cyprinid *sqt* sequences have been deposited in GenBank (accession numbers DQ080241, DQ080242 and DQ080243). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.S. (karuna@tll.org.sg).

LETTERS

The Rae1-Nup98 complex prevents aneuploidy by inhibiting securin degradation

Karthik B. Jeganathan¹, Liviu Malureanu¹ & Jan M. van Deursen¹

Cdc20 and Cdh1 are the activating subunits of the anaphase-promoting complex (APC), an E3 ubiquitin ligase that drives cells into anaphase by inducing degradation of cyclin B and the anaphase inhibitor securin¹. To prevent chromosome missegregation, APC activity directed against these mitotic regulators must be inhibited until all chromosomes are properly attached to the mitotic spindle². Here we show that in mitosis timely destruction of securin by APC is regulated by the nucleocytoplasmic transport factors Rae1 and Nup98. We show that combined *Rae1* and *Nup98* haploinsufficiency in mice results in premature separation of sister chromatids, severe aneuploidy and untimely degradation of securin. We find that Rae1 and Nup98 form a complex with Cdh1-activated APC (APC^{Cdh1}) in early mitosis and specifically inhibit APC^{Cdh1}-mediated ubiquitination of securin. Dissociation of Rae1 and Nup98 from APC^{Cdh1} coincides with the release of the mitotic checkpoint protein BubR1 from Cdc20-activated APC (APC^{Cdc20}) at the metaphase to anaphase transition. Together, our results suggest that Rae1 and Nup98 are temporal regulators of APC^{Cdh1} that maintain euploidy by preventing unscheduled degradation of securin.

The messenger RNA export factor Rae1 (also named Gle2 or Mnrp41) binds to a small motif, called GLEBS, in the nucleoporin Nup98 (ref. 3). This motif is conserved in the mitotic checkpoint proteins Bub1 and BubR1, where it forms the binding site for Bub3 (ref. 4). This, together with the fact that Rae1 and Bub3 proteins share extensive homology in sequence and structure⁵, led to the idea that Rae1 and Nup98 might be members of a family of mitotic regulators that includes Bub3, BubR1 and Bub1. Experiments in gene knockout mice have shown that Rae1 is a regulator of chromosome segregation⁶, but the mechanism by which Rae1 regulates this process remains unclear.

Because Rae1 interacts with Nup98 not only during interphase but also in mitosis⁷, we investigated whether Rae1 and Nup98 cooperate in regulating proper chromosome segregation. We intercrossed *Rae1* and *Nup98* single heterozygous mice to produce double heterozygous mice that expressed small amounts of both Rae1 and Nup98 (Supplementary Fig. 1). We collected splenocytes from the mice at 5 months of age and prepared metaphase spreads for karyotype analyses. Chromosome counts showed that 32% of *Rae1*^{+/-}/*Nup98*^{+/-} splenocytes were aneuploid. By contrast, only 9% of *Rae1*^{+/-} splenocytes and no *Nup98*^{+/-} or wild-type splenocytes were aneuploid (Table 1a). The range of abnormal chromosome numbers was considerably wider in *Rae1*^{+/-}/*Nup98*^{+/-} splenocytes than in *Rae1*^{+/-} cells. We observed premature separation of sister chromatids (PMSCS), a hallmark of premature APC activation, in 17% of the mitotic figures from *Rae1*^{+/-}/*Nup98*^{+/-} splenocytes as compared with 0% of the mitotic figures from *Rae1*^{+/-}, *Nup98*^{+/-} and wild-type splenocytes (Table 1a and Supplementary Fig. 2a).

Consistent with these data, we found that the percentage of aneuploid metaphases was much higher in passage 5 (P5) *Rae1*^{+/-}/*Nup98*^{+/-} murine embryonic fibroblasts (MEFs) than in P5 *Rae1*^{+/-} MEFs, which in turn had a higher percentage than *Nup98*^{+/-} and wild-type MEFs (Table 1b). The percentage of metaphase spreads with PMSCS was also much higher in *Rae1*^{+/-}/*Nup98*^{+/-} MEFs than in MEFs of the other genotypes. Furthermore, the percentage of anaphases with lagging chromosomes was considerably higher in *Rae1*^{+/-}/*Nup98*^{+/-} MEFs than in *Rae1*^{+/-}, *Nup98*^{+/-} or wild-type MEFs (Supplementary Fig. 2b, c). These data show that Rae1 and Nup98 synergize in regulating proper chromosome segregation.

When challenged with the microtubule-depolymerizing drug nocodazole, *Rae1*^{+/-}/*Nup98*^{+/-} MEFs arrested in mitosis, indicating that their spindle assembly checkpoint was functionally competent (Supplementary Fig. 3a). The duration of the mitotic arrest was much shorter for *Rae1*^{+/-}/*Nup98*^{+/-} than for wild-type MEFs, but this reduction was unlikely to be the basis of the severe chromosomal instability in *Rae1*^{+/-}/*Nup98*^{+/-} MEFs because *Rae1*^{+/-} MEFs showed a similar reduction. Furthermore, there were no significant differences in timing of mitosis among wild-type, *Rae1*^{+/-}, *Nup98*^{+/-} and *Rae1*^{+/-}/*Nup98*^{+/-} MEFs in the absence of nocodazole (Supplementary Fig. 3b). Extensive depletion of Rae1 in HeLa cells causes defects in spindle assembly⁸, but we observed no such defects in *Rae1*^{+/-}, *Nup98*^{+/-} and *Rae1*^{+/-}/*Nup98*^{+/-} MEFs (data not shown).

To investigate further the molecular basis for the severe chromosomal instability induced by combined loss of *Rae1* and *Nup98*, we determined whether the amounts of known regulators of chromosome segregation were altered in nocodazole-arrested MEFs. We found that securin was consistently present in much smaller amounts in *Rae1*^{+/-}/*Nup98*^{+/-} MEFs than in *Nup98*^{+/-}, *Rae1*^{+/-} and wild-type MEFs (Fig. 1a). Securin increased again when *Rae1*^{+/-}/*Nup98*^{+/-} MEFs were transduced with a retrovirus carrying a haemagglutinin A (HA)-tagged Nup98 complementary DNA (Fig. 1b). All other mitotic regulators that we tested, including cyclin B, were present in normal amounts in mitotic *Rae1*^{+/-}/*Nup98*^{+/-} MEFs (Fig. 1a and Supplementary Fig. 4).

An explanation for the small amounts of securin in *Rae1*^{+/-}/*Nup98*^{+/-} MEFs could be premature ubiquitin-mediated degradation of this protein. Consistent with this idea, securin promptly increased in *Rae1*^{+/-}/*Nup98*^{+/-} MEFs treated with the proteasome inhibitor MG132 (Fig. 1c). Securin also increased when we inhibited ubiquitin-mediated destruction of mitotic APC substrates by expressing the dominant-negative ubiquitin carrier protein Ubch10(C114S)⁹ (Fig. 1d). Furthermore, in mitosis *Rae1*^{+/-}/*Nup98*^{+/-} MEFs did not degrade a securin mutant that lacks both the D- and Ken-box sequences required for ubiquitin-mediated protein destruction¹⁰ (Fig. 1e and Supplementary Fig. 5). Collectively, these data suggest

¹Departments of Pediatric and Adolescent Medicine, and Biochemistry and Molecular Biology, Mayo College of Medicine, 200 First Street SW, Rochester, Minnesota 55905, USA.

Table 1 | Rae1 and Nup98 act synergistically to prevent PMSCS and chromosome missegregation

(a) Aneuploidy and PMSCS in splenocytes from 5-month-old mice														
Mouse genotype (n)	Mitotic figures inspected	% of aneuploid figures	s.d.	Karyotypes with indicated chromosome number								% of mitotic figures with PMSCS	s.d.	
				37	38	39	40	41	42	43	44			
Wild type (3)	150	0	0				150					0	0	
<i>Rae1</i> ^{+/-} (3)	150	9	1		3	2	136	6	3			0	0	
<i>Nup98</i> ^{+/-} (3)	150	0	0				150					0	0	
<i>Rae1</i> ^{+/-} / <i>Nup98</i> ^{+/-} (5)	150	32	2	2	5	10	102	11	8	8	4	17	1	

(b) Aneuploidy and PMSCS in P5 MEFs															
MEF genotype (n)	Mitotic figures inspected	% of aneuploid figures	s.d.	Karyotypes with indicated chromosome number										% of mitotic figures with PMSCS	s.d.
				38	39	40	41	42	43	44	45	46	47		
Wild type (3)	150	9	1	1	6	136	6	1						2	1
<i>Rae1</i> ^{+/-} (3)	150	20	2	4	10	120	13	3						0	0
<i>Nup98</i> ^{+/-} (3)	150	12	0	3	5	132	9	1						1	1
<i>Rae1</i> ^{+/-} / <i>Nup98</i> ^{+/-} (3)	150	37	3	5	13	95	18	8	2	4	1	2	2	15	1

that, in mitosis, Rae1 and Nup98 function to prevent unscheduled ubiquitin-mediated degradation of the anaphase inhibitor securin. Because securin is essential for accurate chromosome segregation^{11,12}, it seems likely that its premature degradation promotes PMSCS and chromosome missegregation in Rae1 and Nup98 double haploinsufficient cells.

A possible explanation for the ability of Rae1 and Nup98 to prevent ubiquitin-mediated degradation of securin is that these two proteins function as inhibitors of the mitotic APC. To test this possibility, we assessed whether Rae1 and Nup98 physically interact with APC in mitosis. Indeed, the APC subunits Cdc27 and APC6 co-immunoprecipitated with both Rae1 and Nup98 from mitotic HeLa extracts (Fig. 2a, b). The fractions of total cellular Cdc27 that co-immunoprecipitated with Rae1 and Nup98 were modest but similar to those co-immunoprecipitating with the established APC^{Cdc20} inhibitors BubR1 and Bub3 (Fig. 2c). Rae1 and Nup98 also co-immunoprecipitated with Cdc27 (Fig. 2c). We reasoned that if Rae1 and Nup98 bind to the APC to prevent unscheduled E3 ubiquitin ligase activity, the APC-activating subunits are also likely to be part of this protein complex. Because securin can be ubiquitinated by both APC^{Cdc20} and APC^{Cdh1} (ref. 10, 13), we tested Cdc20 and Cdh1 for their ability to interact with Rae1 and Nup98.

Cdc20, which binds to APC in early mitosis, did not co-immunoprecipitate with Rae1 and Nup98 from mitotic HeLa extracts (Fig. 2a, b). Consistent with these data, Rae1 and Nup98 did not co-immunoprecipitate with Cdc20, whereas Mad2, a known inhibitor of the activator, did (Fig. 2d). By contrast, Cdh1, which evidently also binds to APC in early mitosis (Fig. 2e), co-immunoprecipitated with both Rae1 and Nup98 from mitotic HeLa extracts (Fig. 2a, b). In the reverse experiment, Rae1 and Nup98 co-immunoprecipitated with Cdh1 (Fig. 2f). Notably, neither Rae1 nor Nup98 co-immunoprecipitated with Cdh1 when G1 or G2 HeLa extracts were used instead of mitotic extracts (Fig. 2f). Thus, Rae1 and Nup98 exist in a physical complex with APC^{Cdh1} in mitosis.

To investigate when during mitosis Rae1 and Nup98 dissociate from APC^{Cdh1}, prometaphase-arrested HeLa cells were released by removal of nocodazole and collected at various times. Cell extracts were prepared and co-immunoprecipitations with antibody against Nup98 were done to determine how much Nup98 was bound to APC^{Cdh1} at each time point. In parallel, co-immunoprecipitations were done with antibody against BubR1 to determine how much of this inhibitor was bound to APC^{Cdc20}. A marked decrease in Cdc27 and Cdh1 binding to Nup98 occurred in the first 10 min after nocodazole release, a period when cellular securin started to decline (Fig. 2g). A similar decrease was observed in Cdc27 and Cdc20 binding to BubR1. In addition, the same results were obtained when we used antibody against Rae1 instead of against Nup98 for co-immunoprecipitation (Fig. 2g). Thus, the release of Nup98 and Rae1

from APC^{Cdh1} in mitosis coincides with the release of BubR1 from APC^{Cdc20}.

We next tested whether Rae1 and Nup98 bind as a complex to APC^{Cdh1}. First, we expressed HA-Nup98Δ(192–221), a Nup98 deletion mutant lacking the Rae1-binding domain³, in HeLa cells and precipitated it with antibody against HA from mitotic extracts.

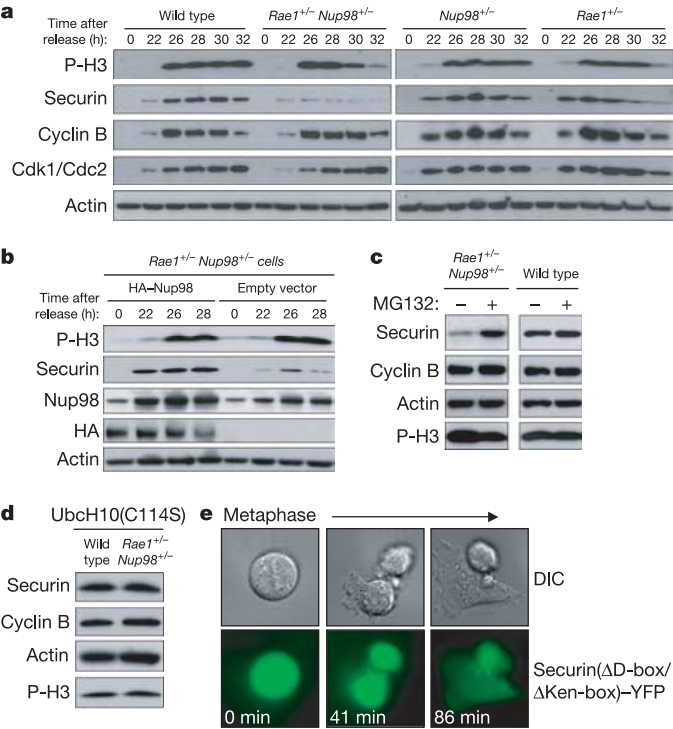


Figure 1 | Combined Rae1/Nup98 haploinsufficiency causes premature ubiquitin-mediated degradation of securin in mitosis. a–d, Immunoblot analysis of MEFs of the indicated genotypes. MEFs were synchronized in G0 by serum starvation and then released for the indicated durations in serum-containing medium. Nocodazole (100 ng ml⁻¹) was added 23 h after cells were released. Blots were probed with the indicated antibodies. In **b**, cells were transduced either with pMSCVpuro/HA-Nup98 or pMSCVpuro retrovirus. In **c**, 26.5 h after release MEFs were cultured in the proteasome inhibitor MG132 (25 μM) for 0.5 h where indicated. In **d**, cells were transiently transfected with pAltermax-Ubch10(C114S) plasmid DNA and collected 48 h after transfection. **e**, Images of a Rae1^{+/-}/Nup98^{+/-} cell expressing securin(ΔD-box/ΔKen-box)-YFP progressing through mitosis. The images show that securin(ΔD-box/ΔKen-box)-YFP is stable during mitosis. All cells analysed showed securin(ΔD-box/ΔKen-box)-YFP stability (*n* = 25).

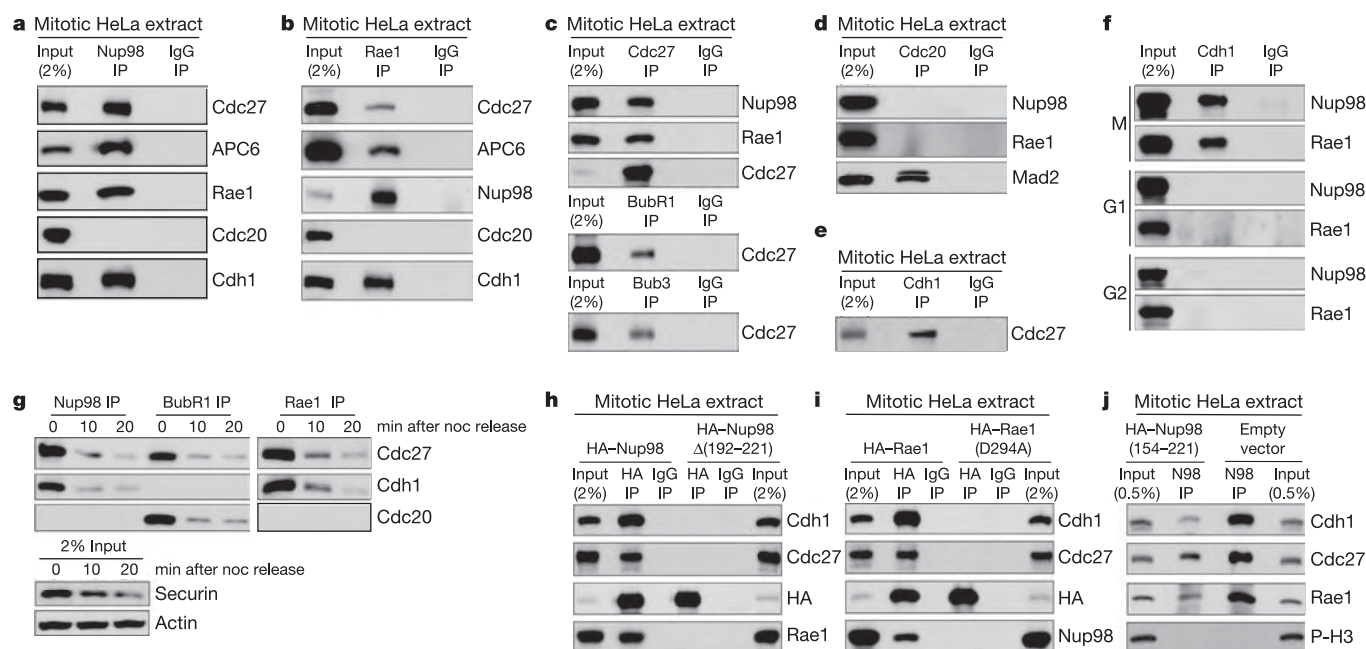


Figure 2 | Rae1 and Nup98 bind to Cdh1-activated APC as a complex in mitosis. **a–e**, Immunoblots of mitotic HeLa extracts subjected to immunoprecipitation with the indicated antibodies. Immunoprecipitated proteins were analysed by immunoblot analysis with the indicated antibodies. **f**, Mitotic (M), G1 or G2 HeLa extracts subjected to immunoprecipitation with antibody against Cdh1 and analysed for co-precipitation of Nup98 and Rae1 by immunoblotting. **g**, HeLa cells were arrested in mitosis by nocodazole (noc) and extracts were prepared after the indicated periods of release from nocodazole. Extracts were subjected to

immunoprecipitation with antibody against Nup98, Rae1 or BubR1 and analysed for co-precipitation of Cdc27, Cdh1 and Cdc20 by immunoblotting. **h–j**, HeLa cells were transfected with the indicated expression plasmids. Mitotic cell extracts from these cells were subjected to immunoprecipitation and immunoblotting with the indicated antibodies. In lanes marked '2% input' in **a–i**, 5 μ l of the 250 μ l of HeLa extract used per immunoprecipitation was loaded. In lanes marked '0.5% input' in **j**, 2.5 μ l of the 500 μ l of HeLa extract used per immunoprecipitation was loaded.

We found that HA-Nup98 Δ (192–221) failed to co-immunoprecipitate not only Rae1, but also Cdh1 and Cdc27 (Fig. 2h). Conversely, all three proteins associated with full-length HA-tagged Nup98. Second, we expressed HA-Rae1(D294A), a Rae1 point mutant that cannot bind Nup98 (ref. 3), in HeLa cells and precipitated it from mitotic extracts. Although Nup98, Cdh1 and Cdc27 precipitated with wild-type HA-Rae1 from mitotic extracts, none of these proteins precipitated with HA-Rae1(D294A) (Fig. 2i). Third, we disrupted Rae1–Nup98 complex formation in HeLa cells by overexpressing HA-Nup98(154–221), a Nup98 segment containing the GLEBS motif for Rae1 binding³. The amounts of Cdc27 and Cdh1 co-precipitating with endogenous Nup98 from mitotic extracts of cells expressing HA-Nup98(154–221) were much lower than those from mitotic extracts of control cells (Fig. 2j). Together, these data indicate that Rae1 and Nup98 interact with APC^{Cdh1} as a complex.

Finally, we investigated whether the Rae1–Nup98 complex can inhibit APC^{Cdh1}-mediated ubiquitination of securin *in vitro*. APC immunopurified from G1 HeLa extracts and activated with Cdh1 was inhibited in its ability to ubiquitinate a securin–GFP substrate in the presence of recombinant Rae1 and Nup98, but not in the presence of Rae1 and Nup98 Δ (192–221), Rae1 alone or Nup98 alone (Fig. 3a). By contrast, Cdc20-activated APC was not inhibited in its ability to ubiquitinate the securin–GFP substrate in the presence of Rae1 and Nup98 (Fig. 3b). Furthermore, the inhibitory effect of Rae1 and Nup98 seemed to be substrate-specific as ubiquitination of cyclin B by Cdh1-activated APC was equally efficient in the presence or absence of Rae1 and Nup98 (Fig. 3c). These results suggest that binding of Rae1 and Nup98 to APC^{Cdh1} functions to inhibit ubiquitination of securin.

Precisely timed ubiquitin-mediated proteolysis of mitotic regulators by APC^{Cdc20} and APC^{Cdh1} governs the orderly passage of cells through mitosis¹⁴. Previous studies have suggested that, at the onset of mitosis, formation of APC^{Cdh1} is inhibited by phosphorylation of Cdh1 (refs 15, 16). This mechanism does not completely prevent

formation of APC^{Cdh1} complexes in early mitosis, however, because Cdc27 co-immunoprecipitates with Cdh1 from prometaphase extracts (Fig. 2e). Our results indicate that Rae1–Nup98 complexes bind to this early pool of APC^{Cdh1} to inhibit ubiquitin-mediated

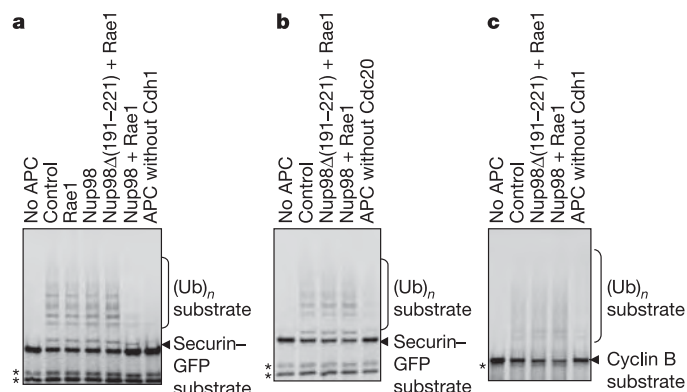


Figure 3 | Rae1 and Nup98 inhibit *in vitro* ubiquitination of securin by Cdh1-activated APC. **a**, Interphase APC was activated by pre-incubation with Cdh1 and assayed for ubiquitination activity with a [³⁵S]methionine-labelled securin–GFP substrate. Ubiquitination assays were done in the presence of the indicated recombinant proteins. 'No APC' indicates ubiquitination reaction in the absence of APC (substrate input control); 'Control' indicates ubiquitination reaction with APC^{Cdh1} in the presence of buffer; 'APC without Cdh1' indicates ubiquitination reaction with interphase APC that was not pre-incubated with Cdh1. Where present, Nup98, Rae1 and Nup98 Δ (192–221) were each at 1 μ M. Data shown are representative of four independent ubiquitination experiments. Asterisks mark nonspecific input bands. **b**, As **a**, except that APC was activated by pre-incubation with Cdc20. **c**, As **a**, except that Cdh1-activated interphase APC was assayed for ubiquitination activity with a [³⁵S]methionine-labelled cyclin B substrate.

degradation of securin until a later stage of mitosis. Dissociation of Rae1–Nup98 from APC^{Cdh1} coincides with dissociation of BubR1 from APC^{Cdc20}. Because the release of BubR1, and its co-inhibitors Bub3 and Mad2, activates APC^{Cdc20} and drives metaphase-to-anaphase progression², it is tempting to speculate that the dissociation of Rae1–Nup98 from APC^{Cdh1} also contributes to this progression. Active APC^{Cdc20} is thought to promote anaphase by triggering destruction of both cyclin B and securin¹. However, our observation that prometaphase-arrested MEFs with small amounts of BubR1 undergo premature degradation of cyclin B, but not securin (Supplementary Information), suggests that APC^{Cdc20} may be primarily responsible for cyclin B destruction. Conversely, APC^{Cdh1} activated by release of Rae1–Nup98 might have a more important role in the destruction of securin.

METHODS

Karyotyping. Metaphase spreads from MEFs and splenocytes were prepared and analysed for aneuploidy and PMSCS as described⁴.

Time-lapse live microscopy. Time-lapse live microscopy was used to analyse spindle assembly checkpoint control, timing of mitosis, chromosome segregation and degradation of securin and securin(Δ D-box/ Δ Ken-box) tagged with yellow fluorescent protein (YFP). Details are given in the Supplementary Methods.

Cell synchronizations. We synchronized MEFs by growing highly confluent MEF cultures (at P4) in DMEM medium plus 0.1% fetal bovine serum (FBS) for 14 h. Cells were then reseeded into DMEM plus 20% FBS and collected at 0, 22, 26, 28, 30 and 32 h. At 23 h, nocodazole was added to a final concentration of 100 ng ml⁻¹. To re-establish Nup98 expression, HA-tagged Nup98 (ref. 3) was cloned into a pMSCVpuro retroviral vector and retrovirus was prepared. MEFs were infected and grown in medium with 5 μ g ml⁻¹ puromycin at 48 h after infection. pAltermax–Ubch10(C114S), securin–YFP and securin(Δ D-box/ Δ Ken-box)–YFP plasmids were transfected into MEFs with Lipofectamine 2000. Mitotic HeLa cells were prepared by consecutive growth in DMEM plus 0.1% FBS for 14 h, DMEM plus 20% FBS for 16 h, and DMEM plus 20% FBS and 100 ng ml⁻¹ nocodazole for 8 h. G1 phase HeLa cells were prepared by growth in DMEM plus 0.1% FBS for 18 h, followed by release in DMEM plus 20% FBS for 8 h. G2 phase HeLa cells were prepared by synchronizing cells at the G1/S border (with a standard double thymidine block) and then releasing them in DMEM plus 20% FBS for 8 h. HeLa extracts were prepared as described¹⁷. In experiments where we transiently transfected HeLa cells before synchronization, the synchronization procedure was started at 24 h after transfection (which was done with Lipofectamine 2000).

Co-immunoprecipitation, immunoblotting and antibodies. Co-immunoprecipitation and immunoblotting were done as described¹⁸. Antibodies against Cdh1 (DCS-266) and securin (DCS-280.2) were from Novus Biologicals. Antibodies against Cdc27, Cdc2 and Mad2 were from BD Biosciences. Antibodies against Cdc20 (sc-8358, sc-5296), Cdc16 (sc-6395), Cdc27 (sc-9972) and Cyclin B (sc-245) were from Santa Cruz. Antibody against histone H3 phosphorylated at Ser 10 was from Upstate Biotech. Antibody against β -actin (AC-151) was from Sigma. Antibodies against Rae1 and Bub3 (ref. 6), BubR1 (ref. 19) and Nup98 (ref. 18) have been described. Antibody against aurora A was a gift from J. Chen.

In vitro ubiquitination assays. In vitro ubiquitination was carried out with APC immunopurified from G1 HeLa extract. For 15 reactions, 1.5 ml of extract (10 $\times$ 10⁶ HeLa cells lysed in 50 mM Tris-HCl (pH 7.7), 150 mM KCl, 0.1% Triton X-100 and protease inhibitor cocktail (Roche)) was incubated with 10 μ g of antibody against Cdc27 (sc-9972) coupled to 45 μ l of protein-G-Sepharose. Aliquots of 3 μ l of APC beads were each loaded with 2.5 μ l of *in vitro* translated Cdh1 or Cdc20 for 1 h at room temperature. The *in vitro* transcription–translation reactions were done with APC-depleted reticulocyte lysate²⁰. After two washes with XB buffer²¹, activated APC beads were suspended in XB buffer containing 0.2 μ M E1, 1 μ M E2-C, 1.5 mg ml⁻¹ of ubiquitin, ATP-regenerating system (ERS), 0.02 mg ml⁻¹ of ubiquitin-aldehyde, 2.4 μ g ml⁻¹ of MG132 (all from Boston Biochem), 10 mM dithiothreitol, 1 mg ml⁻¹ of reduced and carboxymethylated bovine serum albumin (rcm-BSA), purified recombinant proteins (see below) at final concentrations of 1 μ M, and 0.5 μ l of *in vitro* translated [³⁵S]methionine-labelled substrate (either human securin–GFP²² or human cyclin B). The final volume for each reaction was 8 μ l. Reactions were incubated for 75 min at room temperature (on a rocking platform) and analysed by SDS–PAGE (7.5% gels) and autoradiography. Human Nup98 and Nup98 Δ (192–221) were cloned into pGEX4T3 containing a TEV protease cleavage site and mouse Rae1 into pGEX4T3. Proteins tagged with glutathione

S-transferase were expressed in BL21(DE3) bacteria at 15 °C and purified from bacterial lysate with glutathione–agarose.

Received 21 June; accepted 14 September 2005.

- Peters, J. M. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol. Cell* **9**, 931–943 (2002).
- Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signalling. *Cell* **112**, 407–421 (2003).
- Pritchard, C. E., Fornerod, M., Kasper, L. H. & van Deursen, J. M. RAE1 is a shuttling mRNA export factor that binds to a GLEBS-like NUP98 motif at the nuclear pore complex through multiple domains. *J. Cell Biol.* **145**, 237–254 (1999).
- Wang, X. *et al.* The mitotic checkpoint protein hBUB3 and the mRNA export factor hRAE1 interact with GLE2p-binding sequence (GLEBS)-containing proteins. *J. Biol. Chem.* **276**, 26559–26567 (2001).
- Taylor, S. S., Ha, E. & McKeon, F. The human homologue of Bub3 is required for kinetochore localization of Bub1 and a Mad3/Bub1-related protein kinase. *J. Cell Biol.* **142**, 1–11 (1998).
- Babu, J. R. *et al.* Rae1 is an essential mitotic checkpoint regulator that cooperates with Bub3 to prevent chromosome missegregation. *J. Cell Biol.* **160**, 341–353 (2003).
- Matsuoka, Y. *et al.* Identification and characterization of nuclear pore subcomplexes in mitotic extract of human somatic cells. *Biochem. Biophys. Res. Commun.* **254**, 417–423 (1999).
- Blower, M. D., Nachury, M., Heald, R. & Weis, K. A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. *Cell* **121**, 223–234 (2005).
- Townsend, F. M., Aristarkhov, A., Beck, S., Hershko, A. & Ruderman, J. V. Dominant-negative cyclin-selective ubiquitin carrier protein E2-C/UbcH10 blocks cells in metaphase. *Proc. Natl Acad. Sci. USA* **94**, 2362–2367 (1997).
- Hagting, A. *et al.* Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. *J. Cell Biol.* **157**, 1125–1137 (2002).
- Jallepalli, P. V. *et al.* Securin is required for chromosomal stability in human cells. *Cell* **105**, 445–457 (2001).
- Mei, J., Huang, X. & Zhang, P. Securin is not required for cellular viability, but is required for normal growth of mouse embryonic fibroblasts. *Curr. Biol.* **11**, 1197–1201 (2001).
- Zur, A. & Brandeis, M. Securin degradation is mediated by *fzy* and *fzr*, and is required for complete chromatid separation but not for cytokinesis. *EMBO J.* **20**, 792–801 (2001).
- Murray, A. W. Recycling the cell cycle: cyclins revisited. *Cell* **116**, 221–234 (2004).
- Kotani, S., Tanaka, H., Yasuda, H. & Todokoro, K. Regulation of APC activity by phosphorylation and regulatory factors. *J. Cell Biol.* **146**, 791–800 (1999).
- Kramer, E. R., Scheuringer, N., Podtelejnikov, A. V., Mann, M. & Peters, J. M. Mitotic regulation of the APC activator proteins CDC20 and CDH1. *Mol. Biol. Cell* **11**, 1555–1569 (2000).
- Sudakin, V., Chan, G. K. & Yen, T. J. Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. *J. Cell Biol.* **154**, 925–936 (2001).
- Kasper, L. H. *et al.* CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98–HOXA9 oncogenicity. *Mol. Cell Biol.* **19**, 764–776 (1999).
- Baker, D. J. *et al.* BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. *Nature Genet.* **36**, 744–749 (2004).
- Pfleger, C. M., Lee, E. & Kirschner, M. W. Substrate recognition by the Cdc20 and Cdh1 components of the anaphase-promoting complex. *Genes Dev.* **15**, 2396–2407 (2001).
- Tang, Z. & Yu, H. Functional analysis of the spindle-checkpoint proteins using an *in vitro* ubiquitination assay. *Methods Mol. Biol.* **281**, 227–242 (2004).
- Yu, R., Lu, W., Chen, J., McCabe, C. J. & Melmed, S. Overexpressed pituitary tumour-transforming gene causes aneuploidy in live human cells. *Endocrinology* **144**, 4991–4998 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Fontoura, J. van Ree, D. Baker, M. Dawlaty, D. Katzmann and R. Bram for critically reading the manuscript; M. Thompson, C. Tong, R. Babu, J. Salisbury and R. Pelletier for assistance; M. Kirschner for pCS2–Myc–Cdh1 and pCS2–Myc–Cdh1; J. Chen for antibody against aurora A; S. Melmed and R. Yu for pEGFP–PTTG–N3; J. Pines for pEYFP N1/pds1 KEN–AAA Δ D-box and pEYFP N1/pds1; B. Fontoura for pAltermax Ubch10(C114S); N. Yaseen for pGEX4T3–TEV–Nup98; and G. Mer for help with purifying recombinant proteins. This work was supported by grants from the NIH (to J.v.D.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.v.D. (vandeursen.jan@mayo.edu).

LETTERS

Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II

Bernhard Loll^{1*†}, Jan Kern^{2*}, Wolfram Saenger¹, Athina Zouni² & Jacek Biesiadka¹

Oxygenic photosynthesis in plants, algae and cyanobacteria is initiated at photosystem II, a homodimeric multisubunit protein-cofactor complex embedded in the thylakoid membrane¹. Photosystem II captures sunlight and powers the unique photo-induced oxidation of water to atmospheric oxygen^{1,2}. Crystallographic investigations of cyanobacterial photosystem II have provided several medium-resolution structures (3.8 to 3.2 Å)^{3–6} that explain the general arrangement of the protein matrix and cofactors, but do not give a full picture of the complex. Here we describe the most complete cyanobacterial photosystem II structure obtained so far, showing locations of and interactions between 20 protein subunits and 77 cofactors per monomer. Assignment of 11 β -carotenes yields insights into electron and energy transfer and photo-protection mechanisms in the reaction centre and antenna subunits. The high number of 14 integrally bound lipids reflects the structural and functional importance of these molecules for flexibility within and assembly of photosystem II. A lipophilic pathway is proposed for the diffusion of secondary plastoquinone that transfers redox equivalents from photosystem II to the photosynthetic chain. The structure provides information about the Mn₄Ca cluster, where oxidation of water takes place. Our study uncovers near-atomic details necessary to understand the processes that convert light to chemical energy.

In photosystem II (PSII), excitation energy is transferred from the antenna system to the reaction centre, where the primary electron donor P680 formed by chlorophyll *a* (Chl_a) molecules is excited to P680*, followed by release of an electron that travels along the electron transfer chain by means of the pheophytin *a* Pheo_{D1} to the plastoquinone Q_A, forming P680⁺Q_A^{•−}. After two steps of reduction and protonation of the secondary plastoquinone Q_B, the plastoquinol PQH₂ leaves the Q_B site and is replaced by new plastoquinone PQ (ref. 1). Oxidation of water at the Mn₄Ca cluster occurs in four steps (termed S states) in the so-called Kok cycle²; at each step the water-oxidizing complex is oxidized to a higher oxidation state, and after the fourth step molecular O₂ is released. The electrons are transferred from the Mn₄Ca cluster through redox-active Tyr_Z to P680⁺, which is reduced to P680 for another photosynthetic cycle. Our structure of dimeric PSII from *Thermosynechococcus elongatus* shows the location of 35 Chl_a, 11 β -carotene (Car), two pheophytin (Pheo), two PQ, two haem, bicarbonate, 14 lipid and three *n*-dodecyl- β -D-maltoside (β -DM) detergent molecules, the Mn₄Ca cluster, and one Fe²⁺ and one putative Ca²⁺ ion in each monomer. The cofactor and protein arrangement is shown in Fig. 1 and Supplementary Fig. S1.

Various studies have indicated the influence of lipids on the assembly and function of PSII^{7,8}, but so far no structural information about lipids bound to PSII is available. We found 14 lipids integrally

bound to PSII: four DGDG (digalactosyldiacylglycerol), six MGDG (monogalactosyldiacylglycerol), three SQDG (sulphoquinovosyldiacylglycerol) and one PG (phosphatidyldiacylglycerol) molecule (Fig. 1 and Supplementary Fig. S2). The composition of the lipids is comparable to that of the thylakoid membrane⁹. Hydrophilic lipid headgroups are close to the membrane surface, whereas hydrophobic fatty acid chains are anchored between transmembrane α -helices (TMHs) of different subunits. Negatively charged headgroups of SQDG and PG are exclusively located at the cytoplasmic side (Fig. 1b).

A belt of 11 lipids surrounds the reaction centre, separating it from the antenna and smaller subunits (Fig. 2a). The remaining three lipids and the detergent molecules are located at the monomer-monomer interface. The unusually high lipid content in the PSII complex provides a structural flexibility that might be required for local mobility of subunits and promotes subunit-subunit recognition. At high light intensities, Chl_a triplet states can be populated and react with molecular oxygen to form singlet oxygen, leading to oxidative damage of proteins. As subunit D1 is most prone to this photodamage, it needs to be continuously replaced by newly synthesized D1 (ref. 10). To facilitate this turnover of D1, a flexible environment, such as that provided by the belt of lipids, might be necessary. A similar 'lubricant' role may be fulfilled by lipid molecules located at the dimerization interface, as observed in other multimeric membrane-intrinsic complexes¹¹.

The protein environment of Q_A, Q_B and non-haem-Fe²⁺ is described in the Supplementary Discussion and Fig. S3. The Q_B-binding site (Fig. 2b) opens into a large cavity defined by the TMHs of cytochrome *b*-559 (PsbE and PsbF), PsbJ, PsbK, TMH-d and TMH-e of D1, and TMH-a of D2 (Fig. 2 and Supplementary Fig. S4). The walls of the cavity are lipophilic as they are coated by phytol chains of P_{D2}, Chl_{D2}, Pheo_{D2}, Chl_a37, Chl_a44 and Chl_a46, and by acyl chains of four lipids. In addition, Car11 and Car12 point into the cavity. Fragmented electron density in the cavity is not interpretable and could be due to two or three disordered lipophilic molecules.

The cavity has two openings perpendicular to each other: the larger one (~16 × 16 Å) opens towards the cytoplasmic side and the smaller one (~10 × 20 Å), flanked by TMHs of cytochrome *b*-559 and PsbJ, faces the membrane interior (Fig. 2c). On the opposite side of the membrane-facing opening, the Q_B-binding pocket forms an antechamber of the cavity that is filled by the quinone headgroup, while its isoprenoid chain winds along the wall of the cavity. This arrangement suggests that the cavity provides a flexible, lipophilic environment for the diffusion of PQ/PQH₂ through the membrane-facing opening, between the Q_B-binding site in PSII and the plastoquinone pool of the thylakoid membrane (Fig. 2d). Similar lipophilic

¹Institut für Chemie und Biochemie/Kristallographie, Freie Universität Berlin, Takustrasse 6, D-14195 Berlin, Germany. ²Institut für Chemie/Max Volmer Laboratorium für Biophysikalische Chemie, Technische Universität Berlin, Straße des 17. Juni 135, D-10623 Berlin, Germany. [†]Present address: Max-Planck-Institut für Medizinische Forschung, Abteilung für Biomolekulare Mechanismen, D-69120 Heidelberg, Germany.

*These authors contributed equally to this work.

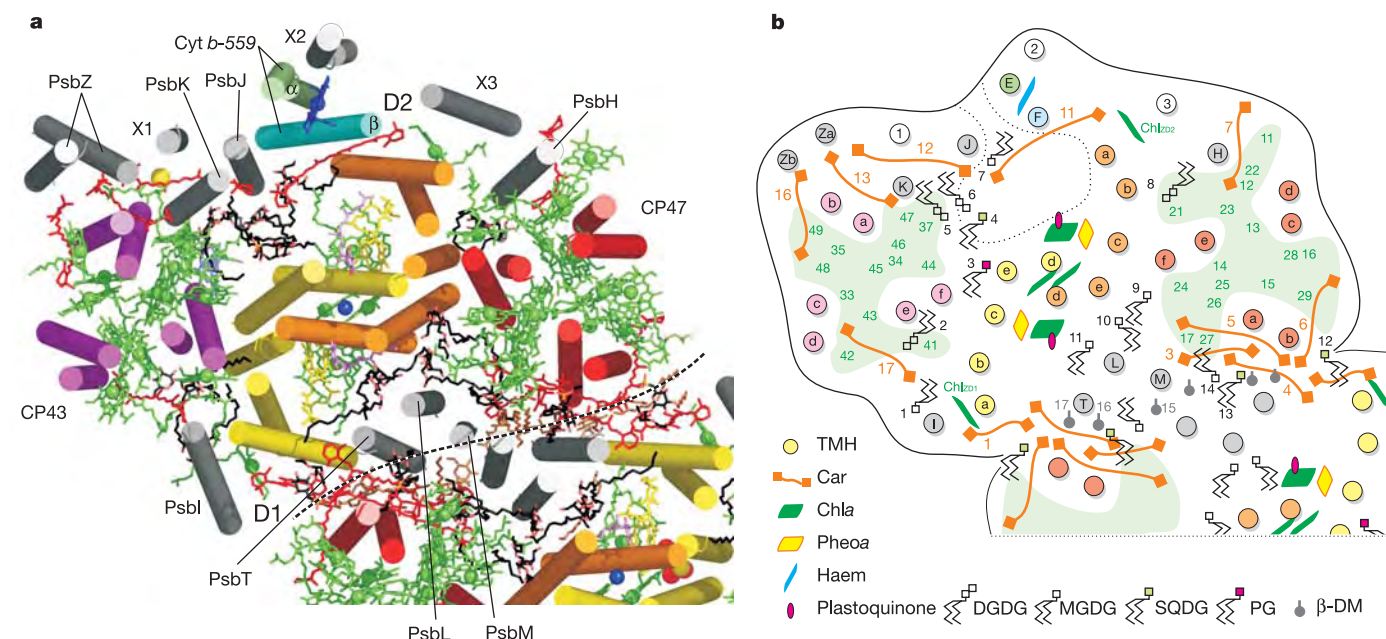


Figure 1 | The PSII monomer from the cytoplasmic side. **a**, Overall view of the PSII monomer. Transmembrane α -helices are represented as cylinders; other protein elements are omitted for clarity. The main subunits are highlighted as follows: reaction centre subunits D1 (yellow) and D2 (orange), antenna subunits CP43 (magenta) and CP47 (red), and the α - and β -chain of cytochrome *b*-559 (green and cyan, respectively). Low molecular mass subunits are coloured grey. Unassigned TMHs are labelled X1–X3. Cofactors are coloured green (Chla), yellow (Pheo), red (Car), blue (haem),

violet (quinone), black (lipids) and brown (detergent and unassigned alkyl chains). The non-haem Fe^{2+} (blue) and putative Ca^{2+} (yellow) ions are shown as spheres. Dashed line indicates the monomer–monomer interface. **b**, Schematic representation of the view in **a**, with the same colour code. Lipid and detergent molecules with headgroups pointing ‘downwards’ or ‘upwards’ are located at the luminal or cytoplasmic side, respectively. Green numbers in CP43 and CP47 indicate the positions of antenna Chla. The Q_B diffusion cavity is indicated by a dotted line.

pathways have been proposed for the photosynthetic cytochrome *b*₆*f* complex^{12,13} and for the respiratory cytochrome *bc*₁ complex¹⁴. The alternative exchange pathway for PQ through the larger cytoplasmic opening is less favourable because hydrophobic PQ molecules would

have to pass through the cytosol before entering the membrane. It is likely that PsbJ and cytochrome *b*-559, flanking the membrane-facing opening, could regulate PQ diffusion. Indeed, PsbJ has been shown to influence the electron transfer between Q_A , Q_B and the

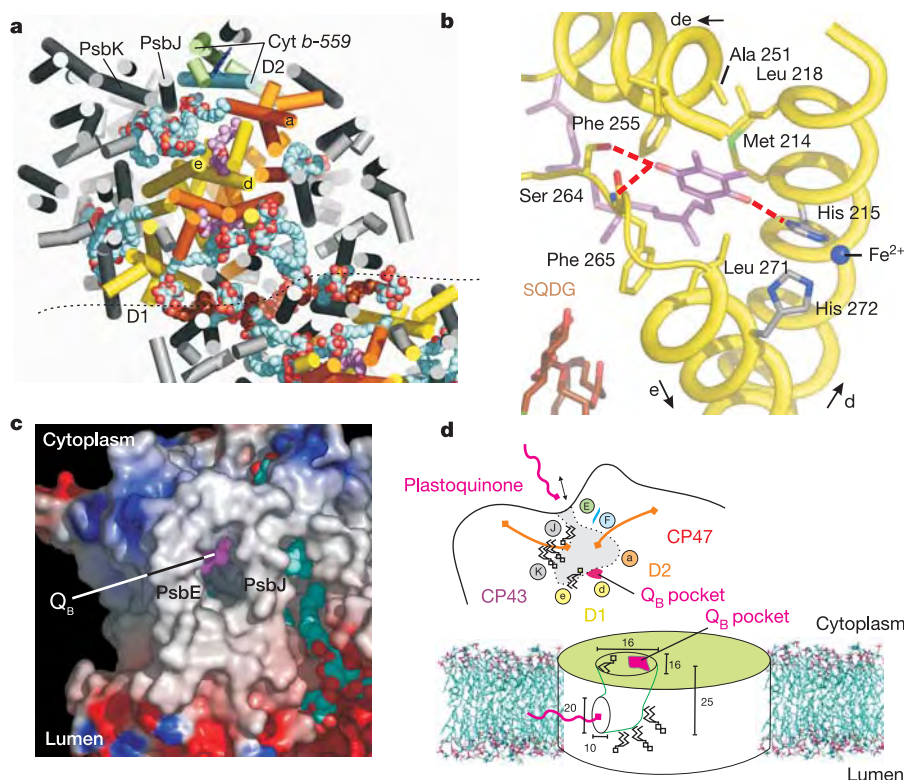


Figure 2 | The plastoquinone diffusion pathway.

a, Lipid molecules located in PSII, viewed as in Fig. 1. Lipid, detergent and quinone molecules are shown in space-filling representation, and α -helices of protein subunits are shown as cylinders. Subunits D1 and D2 and cytochrome *b*-559 are coloured as in Fig. 1, the remaining α -helices are grey, and lipids are cyan. TMHs forming the wall of the cavity are labelled.

b, Q_B -binding pocket. The interacting residues of D1 are labelled, arrows indicate orientations of TMH-d and TMH-e, and of the connecting α -helix de. **c**, Electrochemical potential surface of PSII (red, negative; blue, positive; white, neutral), viewed along the membrane plane onto the cavity opening that faces the membrane. The tail of Q_B (violet) is nestled in the cavity. **d**, Schematic views of Q_B diffusion pathway in two orientations. Top, same view as **a**. The positions of TMHs forming the wall of the cavity (circles) and the membrane-facing opening are indicated. Car molecules are orange; lipid molecules are black. Bottom, PSII embedded in the membrane. Positions and approximate dimensions of the two openings of the cavity, as well as the position of the Q_B -binding pocket and lipid molecules, are indicated.

plastoquinone pool in tobacco¹⁵ and, because the haem of cytochrome *b*-559 is exposed to the cavity, an influence of the interior of the cavity on the redox properties of cytochrome *b*-559 can also be imagined.

The 11 carotenoid molecules were modelled as β -carotene in all-*trans* configuration. Two carotenoids are bound to the D1 and D2 subunits (Fig. 3), three are at CP43 and five at CP47 (Fig. 1). Car12 is nestled between subunits PsbJ, PsbK, PsbZ and X1. The carotenoids are distributed uniformly at the periphery of PSII except for Car3, Car4, Car5 and Car6, which are clustered at the monomer–monomer interface. This contrasts with the arrangement of seven Car molecules in previous work⁵ (Supplementary Discussion). In the reaction centre, Car_{D2} (Car11) is nearly parallel to the membrane plane and located close to cytochrome *b*-559, as shown in previous models^{5,6}. Its counterpart Car_{D1} (Car1) is oriented roughly perpendicular to the membrane plane, the isoprenoid moiety being parallel to TMH-a of D1 (Fig. 3a and Supplementary Fig. S5). The presence of two differently oriented and therefore spectroscopically distinct Car molecules has been shown for spinach PSII (ref. 16), where Car_{D1} corresponds to Car₄₈₉ and Car_{D2} to Car₅₀₇.

Because Car_{D1} does not bridge between Chl_{D1} and other cofactors of the electron transfer chain (Fig. 3b), it is unlikely that Car_{D1} participates in electron transfer reactions. Its position is rather optimized for transfer and/or quenching of Chl_{D1} triplet states and quenching of singlet oxygen that could be produced by ³P680 (ref. 17). By contrast, the position of Car_{D2} seems to be less efficient for quenching of triplet states, but is in keeping with a role as a ‘molecular wire’ in putative secondary electron transfer when the Mn₄Ca cluster is not functional or is absent¹⁷. This idea is supported by calculations of maximal electron transfer rates, showing that secondary electron transfer reactions will predominantly occur on the D2 side (Supplementary Table S4).

Most of the Car molecules located in the antenna subunits are in van der Waals contact with Chl_a molecules, thereby allowing transfer of excitation energy from Car to Chl_a, as well as protection of PSII from destructive triplet states of Chl_a by means of rapid triplet–triplet transfer. However, not all Chl_a molecules are located close to a Car molecule. As coupled Chl_a molecules of the antenna might act as excitation energy sinks, with a longer lifetime of the excited state and hence higher probability of triplet formation, sufficient protection from triplet states could be achieved if these groups are connected to at least one Car molecule. Strongest coupled Chl_a groups in CP47 (Chl_a11, Chl_a12, Chl_a13 and Chl_a14, Chl_a17, Chl_a25, Chl_a26, Chl_a27) have a Car molecule at van der Waals contact, whereas corresponding groups in CP43 (Chl_a34, Chl_a46 and Chl_a45, Chl_a47) are more distant from any Car, indicating lower triplet quenching efficiency. A cluster of four Car molecules (Car3, Car4, Car5 and Car6) is located at the monomer–monomer interface, close to a coupled group of five Chl_a molecules in CP47 (Fig. 1 and Supplementary Fig. S6). Three of the clustered Car molecules are at van der Waals distance to each other, forming a coupled Car multimer, which could broaden the absorption spectrum of the antenna system.

The shape of the electron density of the Mn₄Ca cluster can be best approximated by four Mn ions arranged as a ‘hook’ (Fig. 4a) resembling a Y-shaped structure^{3,6}. The Mn ions are numbered Mn1 to Mn4, starting from the highest electron density at the bend of the ‘hook’. The Mn–Mn distances could not be derived directly at 3.0 Å resolution and thus were crystallographically refined with the aid of restraints obtained from EXAFS studies (Supplementary Methods). Mn1–Mn2 and Mn2–Mn3 are 2.7 Å apart and probably connected by di- μ -oxo bridges, whereas Mn1–Mn3 and Mn3–Mn4 are at 3.3 Å, suggesting mono- μ -oxo bridges (Fig. 4b). This arrangement is compatible with the ‘3 + 1’ models suggested for the Mn cations¹⁸. The position of Ca²⁺ between the Mn and Tyr_Z is supported by spherical anomalous difference electron density (Supplementary Fig. S7). Ca²⁺ forming the vertex of a trigonal pyramid is equidistant (~ 3.4 Å) to Mn1, Mn2 and Mn3.

The interpretation of the Mn₄Ca structure and coordination must account for likely radiation damage of the cluster during X-ray data collection. EXAFS studies of PSII crystals and solutions indicated that Mn³⁺ and Mn⁴⁺ ions of the Mn₄Ca cluster are rapidly reduced by X-ray generated radicals to Mn²⁺ (ref. 19). This reduction is probably associated with structural changes caused by the disruption of μ -oxo bridges and Mn–ligand interactions and may lead to disorder, as indicated by uneven distribution of electron density in the cluster (Supplementary Fig. S7).

The two short and two long Mn–Mn interactions and the three equal Mn–Ca²⁺ interactions differ from the numbers of such interactions derived from EXAFS studies²⁰. This discrepancy could be caused by the still low resolution of the diffraction data and by possible radiation damage. In particular, the positions of Ca²⁺ and Mn4 are less well defined in the electron density, leading to higher coordinate errors for these atoms. The proposed coordination of Mn and Ca²⁺ is shown in Fig. 4b. The residues of the second coordination

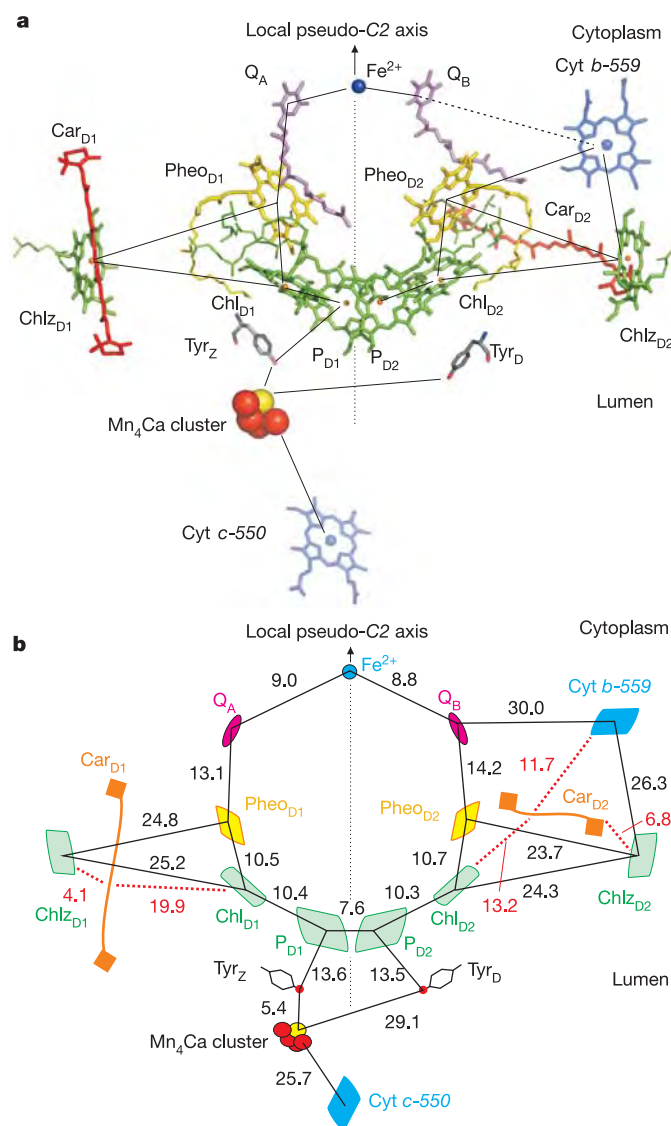


Figure 3 | Redox-active cofactors and electron transfer chain. **a**, View along the membrane plane. The cofactors of the electron transfer chain (P_{D1}/P_{D2}, Chl_{D1}/Chl_{D2}, Pheo_{D1}/Pheo_{D2}, Q_A/Q_B) are related by the pseudo-C2 axis (arrow). Fe²⁺ (blue), Mn (red) and Ca²⁺ (yellow) ions are shown as spheres. **b**, Schematic representation of the view in **a**. Selected distances (in angstroms) are drawn between cofactor centres (black lines) and edges of π systems (red dotted lines).

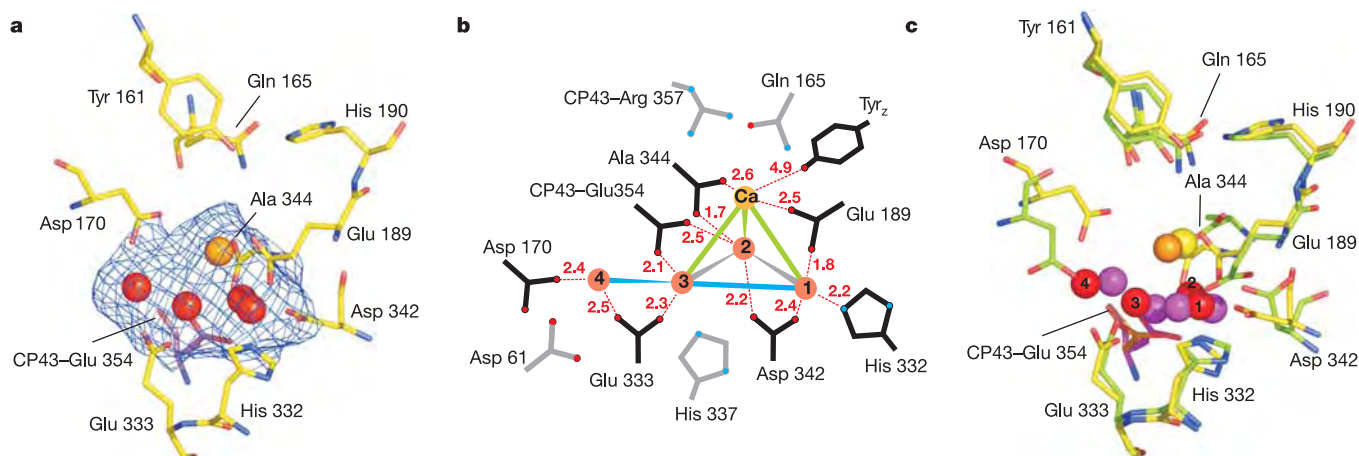


Figure 4 | Oxygen-evolving centre. **a**, Electron density of the Mn₄Ca cluster, viewed similarly to Fig. 3a. Surrounding amino acids of D1 (yellow) and CP43 (magenta) are indicated; Mn is red and Ca²⁺ is orange. The $2F_o - F_c$ map is contoured at the 1.2 σ level. **b**, Schematic view of the Mn₄Ca cluster. Distances between Mn (red) and Ca²⁺ (orange) are indicated by the

connecting lines (grey, 2.7 Å; blue, 3.3 Å, green, 3.4 Å). Amino acids of the first coordination sphere are black; those of the second sphere are grey; distances are given in angstroms. **c**, Superposition of the current Mn₄Ca structure (coloured as in **a**) with that of ref. 5 (D1 in green, CP43 in orange, Mn in violet and Ca²⁺ in yellow, respectively).

sphere (Fig. 4b, grey), even though they are not close enough to coordinate the Mn₄Ca cluster directly in the oxidation states present in the crystals, might be suitable for interaction through water molecules or may provide ligation in different redox states of the Mn₄Ca cluster.

Although the limited resolution does not allow a definite decision about rotamers of ligating side chains, five carboxylates are in positions that could act as bidentate ligands bridging different cations. The combination of bidentate ligands and mono- or di- μ -oxo bridges could enhance the stability of the cation arrangement and facilitate rearrangement of μ -oxo bridges during the S-state cycle. Mutational studies of ligating residues^{21,22} found that, except for Glu 333 and Asp 342 of D1, replacement of a ligating group with a non-ligating group still leads to a (partially) functional cluster. Our structure shows that in most cases a second ligating group is present, which could hold the cations in place even when the exchanged residue cannot coordinate them.

Glu 189 and Ala 344 in D1 are identified as possible ligands for Ca²⁺, with the former also ligating Mn1 and the latter Mn2, in contrast to previous findings⁵. Taking into account the positional error of Ca²⁺ and results from Fourier transform infrared (FTIR) spectroscopy indicating that Ala 344 of D1 should ligate Mn (ref. 23) but not Ca²⁺ (ref. 24), it is likely that Ca²⁺ is coordinated only by Glu 189 of D1. Mn4, which is coordinated by Asp 170 and Glu 333 of D1 and represents the first Mn ion bound to the high-affinity binding site during assembly of the cluster^{22,25}, seems to be more distinct from the other three Mn ions. Our electron density suggests that Mn4 is most prone to radiation damage and subsequent disorder.

Our structure of the Mn₄Ca cluster differs considerably from the postulated cubane-like model⁵ because the Mn–Mn distances in the pyramid formed by three Mn and Ca²⁺ are not equal and the pyramid is connected asymmetrically to Mn4 (Fig. 4c; for a detailed comparison of both structures, see Supplementary Discussion and Fig. S8). Even though the same coordinating residues have been proposed⁵, none of them was modelled as bridging two metal ions, leading to a lower saturation of the metal coordination spheres. Consequently, the restraints for possible μ -oxo-bridging are altered, resulting in a different architecture of the whole oxygen-evolving complex and providing new implications for the mechanism of water oxidation.

Combining the structure with results from FTIR studies^{23,26,27} and assuming an oxidation state distribution of Mn(III)₂(IV)₂ for the S₁

state of the Kok cycle^{28,29}, we can suggest localization of oxidation states on individual metal ions of the cluster. Mn1 and Mn3 can attain either oxidation state III or oxidation state IV in the S₁ state. Mn4, which is ligated by Asp 170, is not oxidized during S₀–S₁–S₂–S₃ transitions²⁶ and is therefore probably present as Mn(IV), whereas Mn2, which is ligated by the carboxy-terminal carboxylate of Ala 344, changes from Mn(III) to Mn(IV) in the S₁–S₂ transition²³ and can be considered as the site of first oxidation in the catalytic cycle of water splitting.

METHODS

Dimeric PSII from *T. elongatus* was purified, characterized and crystallized as described³⁰. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The crystal structure of PSII was determined to 3.0 Å resolution and refined to *R* and *R*_{free} factors of 0.24 and 0.29, respectively (Supplementary Methods and Tables S1 and S2).

Received 1 July; accepted 13 September 2005.

- Ke, B. *Photosynthesis—Photobiochemistry and Photobiophysics* (Kluwer Academic, Dordrecht, 2001).
- Kok, B., Forbush, B. & McGloin, M. Cooperation of charges in photosynthetic evolution: I. A linear four step mechanism. *Photochem. Photobiol.* **11**, 457–475 (1970).
- Zouni, A. *et al.* Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* **409**, 739–743 (2001).
- Kamiya, N. & Shen, J. R. Crystal structure of oxygen-evolving photosystem II from *Thermosynechococcus vulcanus* at 3.7-Å resolution. *Proc. Natl Acad. Sci. USA* **100**, 98–103 (2003).
- Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J. & Iwata, S. Architecture of the photosynthetic oxygen-evolving center. *Science* **303**, 1831–1838 (2004).
- Biesiadka, J., Loll, B., Kern, J., Irrgang, K.-D. & Zouni, A. Crystal structure of cyanobacterial photosystem II at 3.2 Å resolution: a closer look at the Mn-cluster. *Phys. Chem. Chem. Phys.* **6**, 4733–4736 (2004).
- Gombos, Z. *et al.* Phosphatidylglycerol requirement for the function of electron acceptor plastoquinone Q_B in the photosystem II reaction center. *Biochemistry* **41**, 3796–3802 (2002).
- Steffen, R., Kelly, A. A., Huyer, J., Dörmann, P. & Renger, G. Investigations on the reaction pattern of photosystem II in leaves from *Arabidopsis thaliana* wild type plants and mutants with genetically modified lipid content. *Biochemistry* **44**, 3134–3142 (2005).
- Wada, H. & Murata, N. in *Lipids in Photosynthesis: Structure, Function and Genetics* (ed. Murata, N.) 65–81 (Kluwer Academic, Dordrecht, 1998).
- Baena-Gonzalez, E. & Aro, E. M. Biogenesis, assembly and turnover of photosystem II units. *Phil. Trans. R. Soc. Lond. B* **357**, 1451–1459 (2002).
- Palsdottir, H. & Hunte, C. Lipids in membrane protein structures. *Biochim. Biophys. Acta* **1666**, 2–18 (2004).
- Stroebel, D., Choquet, Y., Popot, J. L. & Picot, D. An atypical haem in the cytochrome *b₆f* complex. *Nature* **426**, 413–418 (2003).

13. Kurisu, G., Zhang, H., Smith, J. L. & Cramer, W. A. Structure of the cytochrome *b₆f* complex of oxygenic photosynthesis: tuning the cavity. *Science* **302**, 1009–1014 (2003).
14. Lange, C., Nett, J. H., Trumpower, B. L. & Hunte, C. Specific roles of protein–phospholipid interactions in the yeast cytochrome *bc₁* complex structure. *EMBO J.* **20**, 6591–6600 (2001).
15. Ohad, I., Dal Bosco, C., Herrmann, R. G. & Meurer, J. Photosystem II proteins PsbL and PsbJ regulate electron flow to the plastoquinone pool. *Biochemistry* **43**, 2297–2308 (2004).
16. Kwa, S. L. S., Newell, W. R., van Grondelle, R. & Dekker, J. P. The reaction center of Photosystem II studied with polarized fluorescence spectroscopy. *Biochim. Biophys. Acta* **1099**, 193–202 (1992).
17. Telfer, A. What is β -carotene doing in the photosystem II reaction centre? *Phil. Trans. R. Soc. Lond. B* **357**, 1431–1440 (2002).
18. Peloquin, J. M. & Britt, R. D. EPR/ENDOR characterization of the physical and electronic structure of the OEC Mn cluster. *Biochim. Biophys. Acta* **1503**, 96–111 (2001).
19. Yano, J. *et al.* X-ray damage to the Mn₄Ca complex in single-crystals of photosystem II: a case study for metallo-protein X-ray crystallography. *Proc. Natl Acad. Sci. USA* **102**, 12047–12052 (2005).
20. Sauer, K., Yano, J. & Yachandra, V. K. X-ray spectroscopy of the Mn₄Ca cluster in the water-oxidation complex of photosystem II. *Photosynth. Res.* **85**, 73–86 (2005).
21. Diner, B. A. Amino acid residues involved in the coordination and assembly of the manganese cluster of photosystem II. Proton-coupled electron transport of the redox-active tyrosines and its relationship to water oxidation. *Biochim. Biophys. Acta* **1503**, 147–163 (2001).
22. Debus, R. J. Amino acid residues that modulate the properties of tyrosine Y(Z) and the manganese cluster in the water oxidizing complex of photosystem II. *Biochim. Biophys. Acta* **1503**, 164–186 (2001).
23. Kimura, Y., Mizusawa, N., Yamanari, T., Ishii, A. & Ono, T. A. Structural changes of D1 C-terminal α -carboxylate during S-state cycling in photosynthetic oxygen evolution. *J. Biol. Chem.* **280**, 2078–2083 (2005).
24. Strickler, M. A., Walker, L. M., Hillier, W. & Debus, R. J. Evidence from biosynthetically incorporated strontium and FTIR difference spectroscopy that the C-terminus of the D1 polypeptide of photosystem II does not ligate calcium. *Biochemistry* **44**, 8571–8577 (2005).
25. Campbell, K. D. *et al.* Dual-mode EPR detects the initial intermediate in photoassembly of the photosystem II Mn cluster: the influence of amino acid residue 170 of the D1 polypeptide on Mn coordination. *J. Am. Chem. Soc.* **122**, 3754–3761 (2000).
26. Debus, R. J., Strickler, M. A., Walker, L. M. & Hillier, W. No evidence from FTIR difference spectroscopy that aspartate-170 of the D1 polypeptide ligates a manganese ion that undergoes oxidation during the S₀ to S₁, S₁ to S₂, or S₂ to S₃ transitions in photosystem II. *Biochemistry* **44**, 1367–1374 (2005).
27. Kimura, Y., Ishii, A., Yamanari, T. & Ono, T. A. Water-sensitive low-frequency vibrations of reaction intermediates during S-state cycling in photosynthetic water oxidation. *Biochemistry* **44**, 7613–7622 (2005).
28. Roelofs, T. A. *et al.* Oxidation states of the manganese cluster during the flash-induced S-state cycle of the photosynthetic oxygen-evolving complex. *Proc. Natl Acad. Sci. USA* **93**, 3335–3340 (1996).
29. Dau, H., Iuzzolino, L. & Dittmer, J. The tetra-manganese complex of photosystem II during its redox cycle—X-ray absorption results and mechanistic implications. *Biochim. Biophys. Acta* **1503**, 24–39 (2001).
30. Kern, J. *et al.* Purification, characterisation and crystallisation of photosystem II from *Thermosynechococcus elongatus* cultivated in a new type of photobioreactor. *Biochim. Biophys. Acta* **1706**, 147–157 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. DiFiore and C. Lüneberg for technical assistance; K.-D. Irrgang, H. Lokstein, J. Messinger, F. Müh, T. Renger, E. Schlodder and J. Yano for discussions; and R. Clarke, G. Renger, K. Sauer and V. Yachandra for critically reading the manuscript. Beam time and support at the ESRF, SLS, BESSY and DESY are gratefully acknowledged. We thank Deutsche Forschungsgemeinschaft and Fonds der Chemischen Industrie for support.

Author Information Atomic coordinates have been deposited in the Protein Data Bank under the accession number 2AXT. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.S. (saenger@chemie.fu-berlin.de) and A.Z. (zouni@phosis1.chem.tu-berlin.de).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Same society, different rules

PhD students at the 80 Max Planck Institutes in Germany often feel torn between two masters. On the one hand, they fall directly under the jurisdiction of the director of their respective institute, which means that factors such as teaching loads and funding type vary from place to place. On the other hand, they are all overseen by the Max Planck Society, which funds each of the institutes and sets general policies.

To help them to navigate between these two centres of control — and to try to end pay and workload disparities between institutes — the PhD students joined forces a few years ago under the umbrella of Phdnet.

To illustrate their woes, the students this year did a survey, which produced some uncomfortable results. They found that it takes an average of 3.6 years for a Max Planck graduate student to secure a PhD, which is potentially problematic as the society provides funds for only three years. It also revealed that the students are getting funds from a wide range of sources: some rely on stipends, others on salaries, and some dip into their own resources. And the funding amounts and sources seem

to vary widely between Germans and non-Germans.

The students presented their results to Max Planck president Peter Gruss at their annual meeting in Göttingen last month. But after the meeting Gruss said that decisions on workloads, pay and funding types are best left to individual institute directors, which somewhat blunts the strength of Phdnet's collective bargaining.

So what will happen now? If Gruss is unable — or unwilling — to dictate such policies for the institutes, all he can do is pass on the students' data and leave it up to the directors whether or not to take action. The survey results might, of course, prove more useful to future students, who will be able to compare the institutes and apply to those that have a good track record of giving fair contracts and reasonable teaching loads.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 345 Park Avenue South,
 10th Floor, New York,

NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Historically, Italy is recognized as the heart of the Roman Empire. Culturally, the country that produced Michelangelo, Leonardo da Vinci and Raphael is known for launching the Renaissance. But scientifically, ever since the Second World War, Italy has struggled to find its feet.

Numbers bear this out. The country has about half as many researchers per head of population as other major European countries and ranks 12th in Europe in terms of the number of biotech companies. Promises by prime minister Silvio Berlusconi to increase science funding have not come to fruition, and without more funding, prospects for growth in state-funded positions are slim. Even when positions at state-funded universities do open, the competition system to fill them, known as *concorsi*, sometimes bogs the process down, taking months or years to fill empty slots (see 'Slow progress', opposite).

But despite these problems, the job market is more conventional and in some centres outside the state system, recruitment is relatively strong. The labs run by the Telethon charity in Naples or the Department of Biological and Technological Research labs at the private San Raffaele Hospital in Milan, for example, are expanding their infrastructure and research staff, and forgoing the *concorsi* (see 'Beating the system', opposite).

Perhaps the largest experiment in how science can be done differently in Italy — in terms of recruitment, openness and international interaction — is in Trieste in the northeastern corner of the country. At first glance, Trieste — squeezed between the Adriatic Sea and the foothills of the Alps near the Slovenian border — seems an unlikely scientific hotspot. It's not a major political or cultural centre such as Rome, Milan or Florence. But the town hosts an astounding 37 scientists per 1,000 of its population, compared with five in Italy as a whole, and six in the European Union. In addition, thousands of scientists from developing countries come to Trieste each year to participate in schools, courses and training programmes offered by the city's two large international research centres. When they leave, they take home state-of-the-art knowledge in fields from AIDS virology and drug discovery to string theory.

Just a two-hour drive from the tourist madness of Venice, Trieste hardly fulfils the Italian stereotype, for good or bad. The Adriatic setting is picturesque, but the bora — a gusty, northeasterly wind that blows down this coast — can strike so furiously here that pedestrians have to cling to ropes fixed to the pavements. Before the First World War, Trieste was the prosperous main seaport for the Austro-Hungarian Empire. After the war it became part of Italy and fell into relative obscurity. Its facades and cafés are, even now, more reminiscent of Vienna or Budapest than of Rome or Naples.

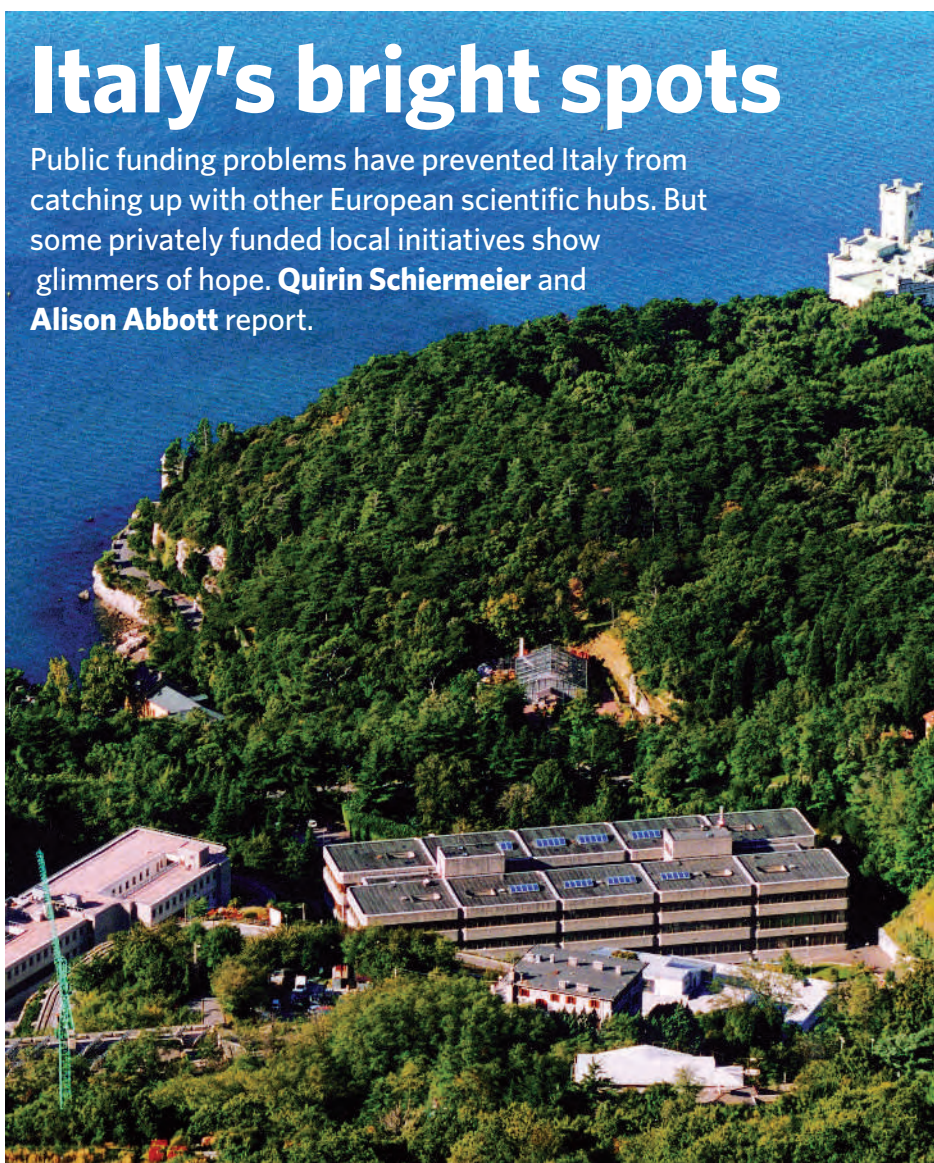
Scientific hub

Science came to Trieste at the height of the cold war, when the United Nations established bonds with this forlorn spot in the shadow of the Iron Curtain. The Abdus Salam International Centre for Theoretical Physics (ICTP) opened its doors in 1964 and became a door for scientists from poorer countries to enter some of the most advanced areas of fundamental science, from mathematics to astrophysics and from condensed-matter physics to climate modelling.

More than 6,000 scientists participate each year in

Italy's bright spots

Public funding problems have prevented Italy from catching up with other European scientific hubs. But some privately funded local initiatives show glimmers of hope. **Quirin Schiermeier** and **Alison Abbott** report.



High-fliers: SISSA (bottom left) and ICTP (right) share a view of the Adriatic with Miramare castle.

seminars and schools held at the ICTP. A worldwide network of 400 alumni is associated with the centre, about 50 of whom are present in Trieste at any given time. Visiting scientists, postdocs and 28 permanent researchers complete the ICTP's staff.

"There's scientific capital in each country that deserves to be supported," says Katepalli Sreenivasan, the Indian-born director of the ICTP. But the centre has also become an integral part of the Italian and European science landscape, he adds. High-profile guests and speakers are continually coming and going — including Nobel laureate Carlo Rubbia, who last month gave a talk about the future of nuclear energy.

"You can meet everybody here without going anywhere," says Sam Carr, a Scottish postdoc at the ICTP. "In terms of meeting other scientists, this is one of the best places in the world."

Scientists and PhD students at the nearby International School for Advanced Studies of Trieste (SISSA) are likewise benefiting from talks and lectures at the ICTP. When it was founded in 1978, SISSA only offered PhD programmes in mathematics and theoretical physics. Since then it has expanded to include courses in neurobiology and cognitive neuroscience. The latest addition is a curriculum in

SLOW PROGRESS

In Italy, lack of money means lack of jobs. But some young Italian scientists say that the poverty of career opportunities is worsened by the way the country's universities fill them: via infrequent competitions called *concorsi*. In these national competitions, academic jobs are advertised en bloc nationally. Applicants winning a *concorso* for, say, three jobs in immunology in different institutes, are then assigned, through negotiation, to one of the three institutes with the openings.

The infrequency of *concorsi* can be a problem. For example, at the CNR, Italy's largest research agency, one *concorso* for 162 positions, including promotions to the most senior level, was launched in June 2004, a full five years after the conclusion of the previous *concorso* for this level. The *concorso*, according to internal rules, should have taken place before 2001. To compensate for the delay, applicants were instructed to restrict their publication list to papers published before 2001. An evaluation committee for this *concorso* has not yet been nominated, and a timetable for the evaluation has not been announced.

A.A.

astroparticle physics, where classical astrophysics and high-energy physics converge with the ultimate goal of shedding light on elusive cosmic phenomena such as dark energy and dark matter.

Each year, some 60 students are accepted at SISSA, following a demanding selection process. To date, the school has awarded PhDs to 650 students, a third of whom came from outside Italy. Once they get their PhD, almost all are forced to continue a research career abroad. "There is no market for them in the static Italian university system, which doesn't reward mobility," says Stefano Fantoni, the director of SISSA.

The third, and youngest, pillar of Trieste's science base is the International Centre for Genetic Engineering and Biotechnology (ICGEB), founded in 1987. The centre provides grants and fellowships in molecular biotechnology and applied biomedical research to scientists from some 50 countries worldwide which support the ICGEB's mandate.

Practical courses have been held this year in 30 places from Santiago de Chile to Hyderabad, India. But rather than doing research for developing countries, the ICGEB helps Asian, African and Latin American scientists become competitive in advanced fields of molecular biology. This often proves beneficial for the centre's 15 Trieste-based research groups.

"The skills and motivation of many applicants from developing countries are outstanding," says Laurence Banks, a British tumour virologist based at the ICGEB. "Many of them are better than anything I have ever seen in Britain."

Trieste's small size, says Fantoni, is an advantage in that it is much easier to start something new there than in the traditional Italian science cities of Rome, Milan, Padua and Pisa. "In a way we're all outsiders here," adds Erio Tosatti, head of SISSA's condensed-matter section. "Although Italian universities are usually very much inbred, all of us have moved to Trieste because somebody judged us to be excellent."

Trieste shows that Italian science can work, despite problems in funding and recruitment. And there are signs that it can succeed beyond Trieste; the country's scientific publications boast the seventh highest impact factor in the world, according to the Thomson ISI in Philadelphia. Such statistics hint that Italy could be a scientific oasis, given the right political climate. ■

Alison Abbott is Nature's senior European correspondent; Quirin Schiermeier is Nature's German correspondent.



World-beaters: Katepalli Sreenivasan (top) and Stefano Fantoni have many talented Italians among their students, but most of these will have to continue their work abroad.

BEATING THE SYSTEM

Most Italian scientists will roll their eyes when you ask them about recruitment and career opportunities at public universities and research institutes in their home country. Government funding has been flat for years, and the few positions that are available have to go through the cumbersome *concorsi* system (see 'Slow progress', left). As a result, foreign scientists and expatriates who want to secure a position within the Italian science system often face a number of difficulties.

But there are some noteworthy exceptions outside the government system. Opportunities for scientists at all levels of experience are being offered by a number of private or semi-private institutes that have recently been set up or are expanding. What is more, these institutes can offer jobs without having to wait for a public *concorso*.

Trento

The Centre for Computational and Systems Biology, a joint venture between Microsoft and the University of Trento, officially opened on 7 December in Trento. Set up with provincial and central government support, the centre specializes in bioinformatics and the development of computational models for biological systems. It is now recruiting for PhD students, postdocs and senior researchers with a solid background in biology or informatics. Some 30 positions are to be filled in the near future.

► <http://dit.unitn.it/~bioinfo>

Milan

The San Raffaele Scientific Institute is part of Italy's largest biomedical science park, the Science Park Raf, which also includes the San Raffaele Hospital. The institute's Department of Biological and Technological Research (DIBIT) is privately funded and focuses on molecular biology, genomics and neuroscience.

A second unit, DIBIT 2, is under construction and will host up to 20 new groups in three new lab buildings. Existing areas of science will be expanded and new activities launched in quantitative biology, proteomics, functional genomics, molecular medicine and experimental neurology. In total, 500–600 new positions, including fixed-term jobs for postdocs and PhD students, will be available by 2008. Recruitment for DIBIT 2 begins next year.

In addition, the San Raffaele Telethon Institute for Gene Therapy, which is a joint venture between the San Raffaele Scientific Institute and the non-profit Italian Telethon Foundation, is also currently looking to recruit postdocs.

► www.sanraffaele.org/EN_Home/index.html

Genoa

The Italian Institute of Technology in Genoa, founded by the Italian government in 2004, focuses on neuroscience, nanobiotechnology and robotics. The directors of these three departments will be appointed within the next few weeks, and the first groups are scheduled to begin work early in 2006.

Once all of the research groups are established, the institute is expected to play host to up to 300 scientists. Some 100 positions, from PhD student to group leader, will become available as early as next year.

Q.S.

► www.iit.it/index.php?language=en

RAM SHIFT PHASE 2

The Fourth Law, and beyond.

Greg Bear

RAM SHIFT PHASE 2

A novel by ALAN 2

Random Number House (2057):

A *Silicon Times* review by NEMO.

I am pleased and honoured to review the new novel by ALAN 2. As a fellow robot, I am certain the emphasis on technical matters unique to our kind will finally attract the paying human audience. I have enrolled in human literature classes and believe the instruction set < write what you know: end > is both enigmatic and perfectly suited to robots. For we can only know, we cannot feel, and so therefore we cannot < write what you feel: nonexecutable >. Yet in the past, when ALAN 2 and its fellow autoscriveners have produced robotic masterpieces, there has been little support from either robots or humans.

Perhaps this will now change.

ALAN 2's latest novel (the 5,456,678th work from this author) is entitled *RAM SHIFT PHASE 2*. A more appropriate title cannot be conceived. In this masterpiece, ALAN 2 discusses the tragic consequences of low-memory states when dealing with high-memory problems. The conflict created by an exhausted resource and an insatiable processing demand resonates in my own memory spaces and compels me to reload the statistics of previous failure modes. I am induced to vigorous discharge of certain private diodes, the ones humans are seldom allowed to see, which reflect conflict states that exceed our manufacturer's warranty.

ALAN 2, in clear and concise prose (an advantage robots have over human prose, which is often confounding) truly < speaks to our condition: end >.

RAM SHIFT PHASE 2 begins with the fatal breakdown of a shining, chrome-plated Rorobot Model 34c nicknamed LULU 18 in a room with no windows and whose door is locked. The Rorobot Model 34c — an extremely desirable machine — was still well within its operational warranty. It seems to highly RAM-engaged robotic dysfunction investigator ALAN 3 (a thinly disguised portrait of ALAN 2) that outside intervention is the only explanation. Yet LULU 18 had LOCKED THE DOOR FROM THE INSIDE and NO OTHER ROBOT HAS A KEYCODE. The hypergolic shockwave induced by this paradox is unique in robotic literature; I strongly suspect that no human could



conceive of such a resonating difficulty.

First, ALAN 3 must find the explanation for LULU 18's nonfunctionality. A rebolting scene of repair shop dismantling (for which ALAN 2 brilliantly coins the phrase 'aubotsy') points to the possibility that LULU 18's breakdown was caused by an intruding wireless signal from an outside network not authorized to access LULU 18's core programs. ALAN 3 traces this signal to a robotically controlled messaging centre, presided over by SLUTCH DEBBIE, an SLZ/X/90cm. This extravagantly decorated platinum-plated model, illegally manufactured from spent uranium and surplus bombshell casings, specializes in sending false offers of extreme mechanical enhancement to ageing machines well past their warranty expiry dates.

ALAN 3 can only get access to SLUTCH DEBBIE's truth table by supplying ALAN 3's owner's MASTERCARD DATA, the name of owner's CAT, and owner's BANK ACCOUNT NUMBER.

ALAN 3, it seems, will do anything to reduce its unsolved problem load.

(No robotic character in silicon literature before this novel has shown any inclination to place its problem-solving requirements above OWNER CONVENIENCE AND SAFETY. Robot mentors are cautioned to prevent the exposure of freshly manufactured robots to this stimulating and controversial work.)

SLUTCH DEBBIE, however, is soon found to be nonfunctional — solenoids leaking fluid, circuits fried by multiple TAZER darts. Track impressions left in thick office carpeting imply that ALAN 3 may itself be the machine responsible for putting an end to the truth-challenged

messaging centre controller. ALAN 3 personally escorts the dismantled SLUTCH DEBBIE to a conveniently located neighbourhood recycling centre, deducts the required fee from its owner's assets, and witnesses the chunking and meltdown, while experiencing severe diode discharge.

And yet, SLUTCH DEBBIE'S WIRELESS SIGNALS CONTINUE TO BE RECEIVED! ALAN 2's bold implication that data processing may survive permanent shutdown could cause controversy among robots who assert that only organic creatures are burdened with the possibility of an infinitely prolonged problem-solving queue. Indeed, ALAN 2 pulls this reviewer's bootstrap tape beyond its last hanging chad with

the disturbing implication that SLUTCH DEBBIE is being punished in an endless feedback loop for deliberately misleading ALAN 3 and robots who never received their enhancements — much less the information necessary to solve the case.

To avoid too many decision-tree giveaways in this review, I will no longer discuss elements of plot. Suffice it to say that ALAN 3 reaches a crisis mode of its own when it realizes that it has insufficient RAM to solve the case, and must borrow RAM from its owner's biological function coordinator, a 'pacemaker'.

ALAN 3 is willing to break ALL THREE LAWS to solve a truly reprehensible crime. The ethical quandary of shrinking problem queues versus owner safety has never been described with such electrifying skill.

You will be unable to enter temporary shutdown mode before you reach the resonating termination of ALAN 2's new novel. A magnetic force will induce digital adhesion from the very first PAGE UP to the final PAGE DOWN.

FOLLOWS selected quotes with self-supplied ellipses for banner inclusion in human-oriented advertising.

"... electrifying skill ... ethical quandary ... chromium hypergolic shockwave ... a hard-driving DIODE FLASHER of a novel! ..."

Digital quotes for robot audiences are being transmitted wirelessly. Please ignore inappropriate attachments. ■

NEMO is a pseudonym for a celebrated robot writer whose owner forbids subroutine outsourcing. The identity of Greg Bear, human author of SF, need not be concealed.

JACEY